

Congress of Paris an anticlimax

This week's diplomatic gathering at Paris has formally confirmed that the Cold War is over, but the future of Europe remains in doubt.

THIS time last year, there was every reason to expect that the gathering of the world's great and good this week in Paris would be a turning point in affairs as significant as the Congress of Vienna which marked the ending of the Napoleonic wars and settled the balance of power in Europe until the Franco-Prussian War of 1870–71. But this year's meeting has turned out differently from that of 1815, and has been an anticlimax. Not that its achievements are negligible. There is, indeed, a treaty on conventional arms in Europe, which has indeed been signed. And the 32 governments represented in Paris have agreed to build on the Helsinki agreements, first negotiated in 1978, which offers ample scope for constructive dialogue in the years ahead.

So why was this week's meeting a disappointment? Part of the explanation is that the proceedings had been too well flagged. Then many of the participants were distracted by happenings elsewhere — the United States with the military situation in the Middle East and the Soviet Union with its survival as an integral entity, for example. But the significant explanation is that none of this week's participants has the faintest idea what the future of Europe will be like. Some of them appear to have been so distracted as not to have cared. And Mr George Bush's puzzlement about "this vision thing", presumably a synonym of "vision", has been infectious.

Yet Europe's future is not nearly as complicated as it has been made to seem this week. Counting Iceland as well as three Baltic states and Albania, Cyprus and Turkey (but supposing that Yugoslavia remains a single state), there are 30 nominally independent states in Europe. Even as things are, they are a diverse lot, some of them more nominally independent than others. Because of Mr Mikhail Gorbachev's continued fudging of the constitution that will define the relationship between the central Soviet state and its constituent republics, it is too soon to know when Russia, Byelorussia, the Ukraine, Azjerbaijan and even Georgia must also be counted separate parts of Europe. So, in an ideal world, the Congress of Paris would have offered a prospectus for a safe and prosperous Europe so beguiling that ditherers on the sidelines would have made up their minds.

Why did it fail? Chiefly because the imagination of the chief actors is still warped by images of the Cold War, whose ending was formalized in Paris. (The treaty on conventional arms, excellent in most respects, is strictly

irrelevant to national concerns over national security now that the Warsaw Pact has ceased to function as a unified military bloc.) But there are more immediate impediments to the general implementation of good sense, among which two stand out: economic disparities, mostly between East and West, and supposed ethnic and cultural differences, which are largely the results of the way in which economically depressed communities are first trapped in old-fashioned and economically low-yielding occupations and then kept there by the mistaken belief that they are protected from an even harsher world by their national sovereignty.

The real Congress of Paris, when it is convened, should start from the notion that Europe is a kind of club to which individuals (rather than their governments) belong, but for whom membership is secured only when their governments conduct themselves in a seemly fashion. The human rights provisions of the Helsinki Final Acts, which would perpetuate existing borders (but not prevent subdivision by consent), which ban discrimination against ethnic and religious minorities and which provide for unfettered emigration (when there is somewhere to go) are a useful starting-point. They need to be complemented by criteria of what constitutes democracy and by provisions, culled from the Treaty of Rome, allowing for the free movement of people in search of work and capital in search of workers. Then it would be possible to look forward to a Europe so quickly made interdependent by more than 500 million people's personal inclinations that this week's remaining difficulties would melt away. Is that not a prize worth aiming at? □

Antarctic wilderness

Mineral exploration should be abandoned, but research should be better co-ordinated and data shared.

THE Antarctic Treaty keeps accreting members, mostly on the strength of governments' eagerness to carry out research programmes of marginal interest on the margins of Antarctica, where the costs and physical difficulties are relatively small. At the meeting beginning this week at Santiago, the treaty members will now embark on a novel exercise — that of going back on a decision reached only two years ago, after endless negotiations, to allow and in the process to regulate the exploration for minerals (in-



cluding office petroleum). Since the Wellington conference at which this agreement was struck in 1988, France and Australia have had a change of heart, and now believe that Antarctica should be held permanently as a wilderness. Other founder-members of the treaty, Britain and the United States for example, seem prepared to fall into step. Other members, Brazil, China and India in particular, will not be so compliant. Where does justice lie?

It makes great good sense that Antarctica should be left much as it is for as long as possible. It is by far the larger of the Earth's two remaining icecaps, and there is every reason for hoping that its integrity will not be threatened in the years ahead by climatic change. For many years, the West Antarctic Ice Shelf has been a particular source of anxiety, mechanically unstable as it may be. If it should slump, become detached and eventually melt, sea level worldwide would increase by a substantial percentage of the 100 metres or so expected from the melting of the whole icecap. But the well-known delicacy of the Antarctic regime is another reason for avoiding unnecessary interference. If this can be agreed during the Santiago meeting, so much the better.

But there are other goals, including the seemingly conduct of the continuing research programme. The usefulness of Antarctic research is amply illustrated by the discovery there, in the Antarctic spring of 1986, of the 'ozone hole', crucial for the proper understanding of the global effects of chlorofluorocarbons on the ozone layer and, indirectly, as greenhouse gases. But much of the research carried out in the Antarctic is too ill-designed to deepen understanding of any kind. Signatories of the Antarctic Treaty cannot be expected to seek approval for their research plans in advance, but it would serve almost as well if all Antarctic researchers were required to present the results of their investigations at regular meetings of kindred spirits. And should there not also be a means by which data gathered by all research programmes should be made generally available within a reasonable time?

A procedure for data-sharing has a direct bearing on the question of Antarctic minerals. The Wellington agreement that commercial exploration should be allowed, with safeguards, sprang from the suspicion that the patches of crustal rock beneath the icecap are a cornucopia of coal, oil and metal ores. Even if the agreement on exploitation is now abandoned, as it should be, the underlying suspicion should be tested. Would it not be strange that Antarctica should be locked up for the rest of time without knowing whether it is full or empty of riches? That is why projects involving drilling into the Antarctic crust, which may be important in understanding the tectonic history of the continent, should be dealt with separately from others. International approval in advance is an obvious need, if only to avert suspicion that they are exploratory ventures in disguise. And there should be arrangements for international access both to the data gathered and even to such physical cores as are recovered. □

Election prospectus

With a British general election due within 19 months, British science should say now what it wants.

Mrs Margaret Thatcher's brush with dissent among her own supporters this week should be a reminder that, after more than a decade of deprivation, the British scientific community should quickly form a coherent view of what it expects from the next government, whatever its political colour. Those who do not ask are unlikely to be rewarded. Here is a skeletal framework for rational expectation:

Pay. For almost 20 years, the salaries of British academics and researchers in the public service have been linked with the rate of general price inflation rather than with the general increase of prosperity. The government's recent exhortation to universities that some people should be paid more than others has cut very little ice so far, and would not in any case diminish the degree to which penury has aggravated researchers' general loss of self-esteem, and of the esteem of other professions.

Research funds. There is no absolute way of telling whether the scale of government support for basic research — £850 million a year in direct payments to the research councils, augmented by an increasingly notional 30 per cent of the recurrent university budget, is sufficient for the need. But there are too few research grants large enough to support internationally competitive research groups even in relatively inexpensive fields. That, as much as the general increase of academics' teaching loads and the distraction of others by extraneous responsibilities, explains why morale is so slow to recover from its nadir (in 1988?). Yet another reorganization is needed, but what?

Students. The research profession differs from most others in the intimacy of its dependence on and solicitude for young people, graduate students in particular. But these are also precisely the young people on whom British hopes for a brighter economic future necessarily depend. There is ample evidence that there are too few of them, especially when allowance is made for the numbers who drift off elsewhere — a tendency that can only be increased by the European single market, still due on 1 January 1993. But there are also disturbing signs that the research profession has acquired such a bad name in recent years that many young bright people are electing to train in other fields. The British educational system has been over-reformed, but yet another round (including a general switch to four-year undergraduate courses) is needed urgently. How will that be done, and who will pay for it? The tendency in the past few months to suppose that students and their families will pay for the expansion needed is an illusion.

The document put out by Save British Science earlier this week (see page 275) pays attention to these same components. Will that organization become the lobbyist the research enterprise needs? □

Chilean funeral for Antarctic minerals pact?

- Minerals treaty in jeopardy
- Negotiations failure would permit mining

Paris, London & Washington

ANTARCTIC researchers should be holding their breath this week as the Antarctic Treaty parties gather in Vina del Mar, Chile for a special consultative meeting to resolve wide differences in opinion on how to control exploitation of Antarctic resources. Depending on how the negotiations proceed, the meeting could end with new rules that could restrict several forms of basic research in the earth sciences. Alternatively, if no agreement can be reached, existing restrictions may be voided and mineral exploitation could begin in earnest.

The meeting was called after the breakdown of eight years of negotiations towards a Convention on the Regulation of Antarctic Mineral Resource Activities (CRAMRA), which attempted to set strict rules for any future mineral exploitation. Although the convention was adopted in June 1988, it was not ratified by the required 16 nations, and so did not enter into force.

Ratification now seems politically impossible, given the refusal of France and Australia to accept any kind of mining activity in Antarctica.

France and Australia will present to the meeting a controversial alternative, co-sponsored by Belgium and Italy, that Antarctica be declared a 'natural reserve' and 'land of science'. The proposal prohibits all mineral resource exploration and exploitation and recommends principles to evaluate the environmental impact of activities, ranging from those to be banned to those with low risk. Formal bodies would be created to oversee the protection of the environment.

The problem for scientists is how the search for mineral resources can be separated from legitimate scientific research. A French foreign ministry spokeswoman says that proposals would be judged "on a case-by-case basis".

But Peter Barker, a geophysicist with the British Antarctic Survey, says he has misgivings about the criteria that would be used to review research proposals. "A pseudo-scientific view of environmental relevance is not a suitable criterion" by which to judge Antarctic research. Rather, he believes that the primary decision should be based on peer review, followed by an independent assessment of the environmental impact.

As examples of work that may suffer, Barker cites seismic reflection, and magnetic and gravitational surveys to investi-

gate the excellent examples of subduction zones found in the Antarctic. Although such research is fundamental to the understanding of plate tectonics, the data would also be useful to the oil industry. Visits by drill ships might also be restricted, he fears, even though drilling

Science Photo Library

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Flying the flag around the South Pole.

Antarctic sediments could provide information on glaciation that would be useful for climate research.

To improve environmental protection, Britain will suggest that a new protocol be added to the existing Antarctic treaty, with specific provisions to deal with wildlife conservation, waste disposal, marine pollution, environmental impact assessments for scientific work, tourism, prospecting and so on.

John Heap, who earned his doctorate in Antarctica and will lead the UK Foreign Office team this week, regards this approach as better than an "overarching idea" on the general principle of environmental protection. 'wilderness park' or 'land of science' proposals are little more than labels, he says, and the British team "doesn't see much virtue in labels. We're concerned about protecting the Antarctic environment".

Environmental groups take a much more positive view of the French-Australian proposal. Simon Lyster, British representative of the World Wide Fund for Nature, concedes that the criteria by which scientific research would be judged under the proposal have not been fully worked out. But he says that "there is no intention that environmental protection should exclude scientific research".

Environmental groups' biggest concern is that the meeting should reach a consensus. James Barnes, an attorney for the Antarctic and Southern Ocean Coalition, which represents some 200 environmental groups and has been granted coveted

observer status for the first time this year, says that the key goal is to keep negotiations from disintegrating. An Antarctic Treaty provision calls for voluntary restraint from prospecting only as long as the members are working towards a timely resolution of the mining issue. If negotiations break down, any party could theoretically begin prospecting unhindered, Barnes says.

The British team still argues that the minerals convention is valuable, but seems grudgingly to accept its demise and are approaching the negotiations "with an open mind" on the minerals issue. Above all, the British team says it will strive to

preserve the consensus that has kept territorial claims to the Antarctic in abeyance since the Antarctic Treaty came into effect in 1961. "Otherwise, the future of scientific cooperation is put at risk" says Heap.

The United States also now regards the proposed mineral convention as dead, helped by pressure from Congress, which in October passed bills call-

ing for a wilderness park and prohibiting US nationals and corporations from mineral resource activities in the Antarctic. The US team has agreed to consider some sort of moratorium on mineral exploitation. Although no length has been specified, State Department officials this month suggested 30 years as a figure that is neither too short to be meaningful nor too long to lock the treaty nations into a decision they might someday regret.

Scientific research within the continent is not without its problems. Barnes says that the Antarctic and Southern Ocean Coalition's immediate goal is a system that "makes more rational decisions on where to build bases." As treaty nation status depends on a member maintaining a full-time scientific base anywhere in the Antarctic, some 10 bases have sprouted up on King George Island, the most accessible and temperate land in the region. "It's easy to establish a claim, but they'll soon run out of science to do there," he says.

UK and US scientists also share the view that there is duplication of work in the Antarctic and that some of the bases achieve little. One solution might be to end the requirement of a year-round base to gain voting rights. Some other criterion, such as possession of a research vessel, could encourage more international collaboration. And on one issue scientists and environmentalists are united: the most immediate threat to the environment is not mineral exploitation but burgeoning tourism.

Peter Coles, Peter Aldhous, Christopher Anderson & Alun Anderson

Researchers fight for a voice

Washington

THE new US federal budget still looks like a victory for the biomedical community with a healthy increase of some 10 per cent granted for the National Institutes of Health (NIH). But as the dust settles, stories are emerging of a full year of behind-the-scenes battling which included not only the usual combat between research lobbyists and congressional staff but also new infighting among scientists themselves. At one point a split between groups backing individual investigators and those representing the universities threatened the increase of nearly \$1,000 million in biomedical research funding in the new appropriations.

The story began early this year, when a group of biologists broke away from the traditional lobbying pack and hired a former congressman named Peter Kyros to be their own Washington representative. That raised eyebrows, especially among the 164 other research and academic groups who, as members of a powerful organization known as the 'ad hoc coalition', carefully craft a science lobbying script each year.

But since then, in a series of questionable moves, the biologists, led by the American Society for Cell Biology (ASCB), have brought the issue of research lobbying to a head, pitting basic researchers against their own institutions.

At the core of the dispute is the way science is represented in Washington. The university groups, especially the Association of American Universities (AAU) and the American Association of Medical Colleges (AAMC), have proved to be remarkably effective in lobbying for more money for research. But in the process, the associations also stressed the need for "generational equity". What that means, the academic lobbyists say, is that new funds should be shared between grants, and money for new facilities and "infrastructure" — provisions for the next generation of researchers.

But in the opinion of the bench scientists, who in 1990 faced the first real decrease in new and competing grants from NIH in recent history, there may not be another generation of researchers without new grants.

Concerned that their message was not getting across, the cell biologists refused to join the *ad hoc* coalition unless they were given a seat on the steering committee where they could ensure that their concerns were being addressed. Denied that, they took matters into their own hands, unleashing Kyros to impress on Congress the value of research by individual investigators.

One of Kyros's first moves was to have several biologists testify at a hearing in

favour of more 'RO-1' grants, the basic individual NIH award. Noting that in 1990, new and competing RO-1s fell to 4,600 — nearly 2,000 less than two years previously — the scientists asked for \$200 million in new funding, enough (assuming an average grant size of \$200,000) to support 1,000 new awards. All this would have been fine, had not Congress already been contemplating an increase of about five times as much.

"It's like going to your richest donor and asking him to buy you lunch", is how AAU president Robert Rosenzweig describes the hearing. One staff member of the appropriations committee recalls: "In a scarce environment sometimes we take the lowest bidder, and they immediately became the lowest bidder". Although Congress eventually got straightened out, there was real concern at the time that the entire research lobbying process had been adversely affected.

Then last month, Kyros helped to set up a Biomedical Research Caucus — an informal group of some 33 Members of Congress who would meet researchers regularly to talk about issues of mutual interest. Unfortunately, the caucus was widely (if perhaps wrongly) perceived as a back-door lobbying effort that undermined the *ad hoc* coalition (see *Nature* 347, 505; 11 October 1990).

The dispute is symptomatic of the increasing division between the 'big science' and 'little science' camps. "There is a real tension building here", presidential science adviser D. Allan Bromley said in an interview earlier this month. "I hope [the scientists] can draw on some innate statesmanship, realizing that it is never an effective technique to blackball competing researchers in lobbying Congress." Direct contact with Congress is important, adds former NIH director James Wyngaarden, "but I'm not sure the best way to do it is with a whole bunch of splinter groups. That kind of behavior just cancels out." As of last week, there was still a good deal of bad blood between basic biomedical researchers and the academic community over the priorities of Washington lobbying. At a meeting in San Francisco earlier this month, ASCB officials decided to retain Kyros and go it alone for now, postponing any decision on whether to join the *adhoc* coalition. Officials from both sides describe the dispute as an unfortunate "breakdown of communications". But through the smoke, they say they can already see hints of raised sensitivities in both groups towards the needs of individual investigators as well as institutions. That alone, says one scarred veteran of this summer's lobby wars, might someday make the whole affair "almost worth it". **Christopher Anderson**

Saddam to the rescue?

Boston

IN A bizarre twist to the decades-long battle over the nuclear plant at Shoreham, New Hampshire, US Department of Energy (DoE) officials are now claiming that the situation in the Middle East may bring the plant back into operation.

The fate of the plant appeared sealed in July 1989 when shareholders voted to sell the completed-but-unopened plant to the state for one dollar. Under the agreement, the state would then be responsible for decommissioning the plant. But at the time, James D. Watkins, the Energy Secretary, decried the closure as "utterly irresponsible" and pledged to make the plant's operation a personal priority.

With events in the Middle East providing the opportunity, DoE has now filed a legal brief to the Nuclear Regulatory Commission (NRC) arguing that before the decommissioning process can begin, NRC must once again reconsider the option of operating the Shoreham plant. According to the brief, the original congressional act that split the DoE from the Atomic Energy Commission enables the Energy Secretary to "order the operation of an electrical generating facility whenever the Secretary determines that there is an energy emergency . . .".

Ironically, that argument puts DoE at odds not only with New York state officials and local residents (the overwhelming majority of whom back the plant's decommissioning), but also with the Long Island Light Company (LILCO) that built Shoreham in the first place. Having sold the plant to the state, LILCO is reluctant to re-open the issue of the plant's operation.

DoE officials last week refused to comment on the brief, saying that it speaks for itself. Clearly, though, the department's contention that it could force Shoreham to open is easier to argue than it would be to implement. **Seth Shulman**

CNRS

NMR turns gold

Paris

THE Centre National de la Recherche Scientifique (CNRS), the largest basic research organization in France, has awarded its 1990 Gold Medal to Professor Marc Julia. Julia is director of the chemistry department of the Ecole Normale Supérieure in Paris and is head of the CNRS 'molecular activation unit'. A member of the Académie des Sciences since 1977, Julia has pioneered new techniques in therapeutic chemistry, including the use of nuclear magnetic resonance. **P.C.**

FRENCH EDUCATION

Back on the streets

Paris

THE reform of education has been the national priority for the French socialist government since the 1988 elections, but the failure to bring about speedy change is provoking protests that seriously challenge the leadership. The problem, which last week produced ugly scenes as hooligans 'pirated' demonstrations by thousands of high-school pupils, is that the government's ambitious reforms are long-term while the problems of education are immediate.

The education minister, Lionel Jospin, has had his hands full since the new term began in October. University reforms proposed in an emergency plan last year involve building much-needed new teaching facilities, creating 1,500 new lecturer-researcher posts, increasing the number of school-leavers who get university places, raising postgraduate grants and doubling the number of PhDs within four or five years. But these reforms will take time to take effect. More than 200,000 square metres of new buildings have been opened, but student intake has increased by only 8 per cent, and there are not enough takers for the new lectureships. Students found themselves faced with much the same overcrowded lecture theatres as last year.

In 1986, under the previous Chirac government, university students took to

the streets to protest against unpopular university reforms. Last week, for the first time, it was the turn of lycée pupils. More than 250,000 from all over France marched from Bastille to the Champs Élysées two days in succession, protesting that reforms were proceeding too slowly. High on their list of priorities are urgent improvements to dilapidated school buildings, more teachers and a return to order. Suburban schools around Paris are frequent targets for gang violence. Ironically, the same gangs took advantage of the demonstrations last week to pillage shops and pelt the security forces, who were under orders not to retaliate.

With a new form of taxation being pushed through parliament, the wave of unrest has taken the spotlight off Jospin and onto the Prime Minister, Michel Rocard. President Mitterrand is now thought to want to dump Rocard. Last week, he sided with the lycée pupils and told Rocard to "do something". Although Rocard immediately found an extra FF4,000 million (\$800 million) for secondary education, this will increase the budget deficit.

Profiting from this mid-term confusion, the opposition challenged the leadership with a vote of no confidence on 19 November, but Rocard and his government survived.

Peter Coles

BRAZIL

Last hope for Atlantic forest?

Paris & São Paulo

IN what may be the last chance to save Brazil's Mata Atlantica coastal rain forest from disappearing, the Brazilian Institute for the Environment and Renewable Natural Resources (IBAMA) wants to turn part of it into a world conservation area. Tania Munhoz, president of IBAMA, was in Paris last week to present the scheme as a candidate for the United Nations Educational, Scientific and Cultural Organisation (UNESCO) Man and Biosphere (MAB) programme. MAB was holding its eleventh international coordinating council meeting at UNESCO's Paris headquarters.

So far, MAB has helped to set up 285 such 'biosphere reserves' in 72 countries. Once stretching over 850,000 square kilometres (10 per cent of the country), the forest has been reduced to about 3 per cent of its former size. Gnawed into by the urban sprawl of cities including Rio de Janeiro and São Paulo, and threatened by fires that spread from neighbouring farm land, the Mata Atlantica has become the second most endangered forest in the world after the rain forest of Madagascar.

IBAMA wants to designate two zones

of the Mata Atlantica as MAB biosphere reserves. One is a 3,500 hectare complex — mostly tropical rain forest — around Rio de Janeiro, including the Tijuca national park and the 137 hectare botanical gardens. The second zone is on the border between the states of São Paulo and Paraná and comprises 22 existing conservation areas. Primates, including the rare lion tamarin, armadillos, several species of endangered birds and new-world cats live in the coastal forest.

Munhoz says that designation of the area as a MAB biosphere reserve will give the regions "a chance of being better conserved and will accelerate the process of protecting the ecosystem". There is little direct financial benefit to be gained from inclusion within MAB. The whole programme receives only \$1.1 million from UNESCO's budget, topped by a further \$6 million from 'extra-budgetary' sources. But IBAMA has been awarded a grant of \$20 million over three years from the World Bank for the Mata Atlantica project and already plans to spend 2,000 million cruzeiros this year on conservation work in the forest.

Peter Coles & Ricardo Bonalume

SCIENCE EDUCATION

Texas votes for evolution

Boston

STUDENTS at Texas public elementary and high schools will now use biology textbooks that teach the theory of evolution unimpeded by creationist views, thanks to a recent vote by the state's Board of Education. The board's approval of new texts for grades 1-12 will determine the choice of textbooks for the next six years and marks an important watershed for the teaching of evolution in public schools in the United States.

Until now, proponents of creationism in Texas have successfully limited coverage of evolution in biology textbooks, favouring instead a view of human origins based on a literal interpretation of the Bible. Because Texas is one of the largest centralized purchasers of textbooks in the United States, its selection process has influenced publishers and, in turn, science curricula throughout the United States.

The board's move to approve new textbooks follows the decision it made last year to require that explanation of the theory of evolution be incorporated into biology textbooks. Previously creationists had succeeded in limiting description of the theory in many of the state's schoolbooks or had required that it be taught only in conjunction with competing creationist views.

Many opponents of creationism, such as Thomas Jukes, a biophysicist at the University of California at Berkeley, hailed the news as the culmination of "a long hard fight" on the part of many participants to "try to make people in



Texas realize that evolution exists". Jukes's colleague, biologist Kevin Padian, agreed, but cautioned that the threat of pseudoscientific views encroaching on science curricula is still very real. Padian urges fellow biologists and other researchers to play a greater part in the way science is taught at the elementary and high-school levels.

Seth Shulman

Pressure for quick release

Washington

IN the wake of yet another round of accusations that US health officials have delayed informing the public of a promising AIDS treatment, the National Institutes of Health (NIH) are moving to establish a set of standards for the quick release of important health-related research findings before they are published in a scientific journal.

Faced with epidemics such as that of AIDS, with its vocal and scientifically literate advocacy groups, NIH officials have been under increasing pressure to short-circuit the traditional journal-based process of disseminating scientific information and instead issue immediate warnings to physicians and the public.

Although a procedure for such release was set up in the mid-1980s to cover advances in cancer research, it has been used sparingly and has been opposed by some researchers, who fear that their chances of publication in a leading journal will be imperilled by early release of their data.

Last week, the sensitivity of the issue was made painfully clear when the *New York Times*, in a front-page article, accused NIH of delaying for five months the release of the results of a study showing that steroids appear to be effective in treating *Pneumocystis carinii* pneumonia, a common and often fatal infection of people with AIDS.

The article said that members of an expert panel had decided that a public announcement should be made, but had spent from May to October of this year

negotiating the wording and timing of that announcement so as not to preclude publication in the *New England Journal of Medicine* (NEJM).

NIH quickly contradicted much of the *New York Times* article. Among other things, a NIH statement noted that the results of the steroid trial had already been widely reported, and had even appeared in a front-page *Los Angeles Times* article in June and that NEJM had agreed to allow an early release of the data. But NIH officials also admit that the controversy points to an unresolved problem in reporting key clinical trials.

Clinical alerts

Samuel Broder, director of NIH's National Cancer Institute (NCI), says that although NCI has a procedure for early notification, known as clinical alerts or updates, it has been used only twice in two years, in part because of criticism from the scientific community.

"When [former NCI director Vincent] DeVita issued the first alert [a breast cancer therapy notice in May 1988], he caught unshirted hell", Broder recalls. When Broder himself decided last year that a new colon cancer treatment called for another early announcement, scientists reacted nearly as badly. "Shirted hell", Broder says, may have been an improvement, but not much of one.

The problem, Broder says, is that many researchers are uncomfortable both with the idea of science by press-release, and the idea of "central bureaucratic author-

ity dictating the release of their work".

NIH officials also accept that there are risks in the circulation of clinical alerts. Broder notes the sobering possibility that the information may be wrong. With an official NIH notice in hand, physicians may well turn immediately to the new treatment, even though nuances that would be clear in a journal article might not be obvious in an abbreviated clinical alert. If further investigation reveals error in the research, "you've got a disaster", he says. "With a government stamp on the information, people tend to view it more uncritically than they would a journal article", says Anthony Fauci, director of the NIH National Institute of Allergy and Infectious Diseases.

Despite these risks, NIH have decided that early release is unavoidable in cases of significant advances that have broad implications for public health. Fauci says the risks can best be reduced by making important data public under the supervision of scientists, health officials and scientific journals. A meeting between these groups has been arranged for 15 January at NIH to establish a procedure for such releases. A critical element, Fauci says, will be to establish a consensus on the issue, both among scientists and the journals.

A sampling of the opinions of top journals suggests this may be easier said than done. Policies on early release of data vary (see below), ranging from absolute prohibition to guaranteed leniency.

Electronic solution

NCI already has the means to make primary data on which an alert is based available to physicians in an on-line electronic database (the PDQ system). But NIH officials acknowledge that many doctors are still unfamiliar with computer telecommunications. They have also suggested that journals maintain on-line databases of their own, to make data available in the period between the acceptance of research articles and their publication. That, too, is likely to find some resistance. Although several journals are considering electronic versions, few have found a way to make on-line access pay for itself.

The need for a uniform policy on eventual publication is also plain. As things are, researchers can rarely predict what might happen to their papers should NIH decide on a clinical alert.

Jerome Groopman, a Harvard University AIDS researcher, says a clear policy is essential, but insists that the direction has to come from the scientific community, not just journal editors and NIH officials. "It's our responsibility to establish our own ethical norms." January's NIH meeting may be a good opportunity to put that to the test.

Christopher Anderson

What the journals are saying

ARNOLD Relman, editor of the *New England Journal of Medicine*, says he will always consider allowing the early release of results with important public health implications. His journal also permits the discussion of results at scientific meetings, and even if it is then "picked up by the press, we would not let that get in the way of publication". But Relman says that "we warn researchers to be cautious at press conferences lest they give away more details than they gave at the meeting, or turn over their manuscript", which could affect eventual publication.

At the other end of the spectrum is *Cell*, which publishes mostly basic, rather than clinical, research. The editor, Benjamin Lewin, says *Cell* has "an absolute ban on any release prior to the release date of the journal". But he says that, with a 'fast-track' review process, *Cell* can publish an article in as little as two weeks after submission. "If there is a threat of a leak — anything that may disturb the natural order of [scientific publishing] — we may

consent to speed up the review process", he says.

This journal and *Science* occupy something of a middle ground. *Nature* will consider allowing researchers to release results once an article has been accepted for publication, but only after negotiation with the authors, according to the editor, John Maddox. Daniel Koshland, the editor of *Science*, takes the same view and recalls several occasions when *Science* has agreed to the early release of papers of medical importance.

It seems that the more clinical a journal, the more tolerant it is of early release. But journals earn their reputations by publishing more new and interesting work than their competitors. Few would tolerate regularly losing their thunder to NIH alerts and press conferences. Fauci suggests that a solution may lie in an independent panel that would review results and, with the agreement of editors, select for early release only those findings most likely to save lives immediately. C.A.

Betting on an inexact science

San Francisco & Tokyo

DESPITE all the uncertainties of earthquake prediction, the state of California this month took the daring step of producing a detailed plan for warning the public of imminent earthquakes. The plan is the first to be developed in the United States, although an early warning system based on the advice of expert seismologists has been operating in Japan for more than ten years. Issued by the Governor of California's Office of Emergency Services, the 140-page *Short-Term Earthquake Prediction Response Plan* lays out how to let millions of citizens know of an increase in the chance of an earthquake in the following 72 hours, and what specific safety precautions communities should take.

The new system provides the best use of current scientific knowledge for public policy purposes, says James Davis, the California state geologist and chairman of the California Earthquake Prediction Evaluation Council (CEPEC). Although some California seismologists say they feel uncomfortable at being forced to discuss the technicalities of their field in public, they generally accept the plan as a reasonable bridge between an inexact science and public policy. But they continue to stress that they are unable to predict earthquakes. Rather, scientists can identify only heightened probabilities of damaging earthquakes in the coming few hours or days because the dataset of clear short-term warning signals is still tiny.

According to the state's plan, short-term predictions of increased seismic activity, which would almost certainly come from the US Geological Survey, are evaluated by CEPEC. The council, consisting of ten seismologists and geologists, convenes quickly, probably by conference call, and determines whether the observations merit a public warning from the Office of Emergency Services. The warnings will allow city and county governments to take precautions such as activating emergency shelters, alerting reserve emergency workers and evacuating high-risk buildings.

In addition, CEPEC has pre-approved statements that are to be issued in case of various earthquake events in the heavily monitored region of the San Andreas fault near Parkfield in central California. A moderate earthquake is expected there very soon, judging from past seismic patterns. A similar pre-approval warning

system is being developed for the southern part of the San Andreas fault.

In a less formal manner, the state has actually issued five earthquake advisories since 1985, including two for the Loma Prieta region of the San Andreas fault, the site of the October 1989 earthquake. However, the second of those five-day advisories came two months before the magnitude 7.1 earthquake, much too early to help.

Although the United States has a National Earthquake Prediction Evaluation Council, that organization focuses on longer-term forecasts for regions throughout the country. Japan, however, has a

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

After-effects of earthquakes — predicting the unpredictable.

complex national early warning plan in place. In the event of possible earthquake precursor activity, the Japan Meteorological Agency (JMA) whisks a committee of six experts to its Earthquake Assessment Committee Room in Tokyo within one hour. Members of the committee carry pocket pagers at all times and are not allowed to go out of pager range without informing the agency.

If a major earthquake is deemed imminent by the committee, the director general of the JMA flashes a message to the office of the prime minister, who then issues a warning to the nation. The Japanese then prepare emergency supplies, make buildings ready, and take cover or evacuate, according to plans rehearsed in annual earthquake drills.

Although the committee has met several times for practice, it has yet to issue an actual earthquake warning. And at least one Tokyo University geophysicist suspects that the aim of the whole complex warning system is to diffuse responsibility among as many scientists, bureaucrats and politicians as possible in case it does not work. All seismologists know that earthquakes can seldom be predicted, he says, but they cannot say that too loudly for fear that the ¥1,800 million (\$14 million) a year in funds for earthquake prediction would dry up.

Elizabeth Schaefer & David Swinbanks

Don't look now . . .

Washington

AFTER a summer of US Department of Energy (DoE) press conferences designed to detail every step of the Superconducting Super Collider (SSC) project's redesign, DoE deputy secretary Henson Moore finally let the bad news slip out last week at a poorly attended signing ceremony in Texas. The latest estimate is that SSC will cost \$8,244 million and take until 1999 to complete, a dramatic change from Moore's brave declaration of three months ago that "we're going to hold the contractor to \$7.8 billion and to finish by 1998" (*Nature* 356,401; 2 August 1990).

What made Moore swallow his words, says a DoE spokesman, is that "the [DoE] review groups said that was too aggressive a funding profile", given congressional budget restraints. All of this, however, is news to Congress. Although last month's budget language directs DoE to serve up an SSC cost estimate "immediately", DoE now says that the White House will not let it report formally the new figures until after the president's 1992 budget request in January. Congress is unlikely to take the snub kindly. "We asked for a number and we didn't get one. They've got to come to us for another budget next year", an appropriations staff member hints darkly.

Christopher Anderson

Exiled writer refuses post

Cape Town

THE University of the Western Cape (UWC)'s vice-rector-designate, Professor Njabulo Ndebele, last week informed the university that he will not now be taking up the post. The appointment of Ndebele, who has lived in exile as the president of the Congress of South African Writers created an uproar in April when the government refused to grant him a work permit (see *Nature* 344,581 & 345,284; 1990). In a statement released last week, the Rector of UWC, Jakes Gerwel, criticized the government, saying that he had no doubt that it had contributed to pressures on Ndebele. Ndebele still intends to return to South Africa, but is now seeking employment in Johannesburg.

Michael Cherry

Happy ending?

London

THE long-running dispute at the Natural History Museum in London may have been resolved now that all the researchers threatened with redundancy have been found alternative employment within the museum. A spokesman from the scientists' trade union, the Institution of Professionals, Managers and Specialists (IPMS) intimated last week that such a situation would be seen as an "acceptable olive branch".

Henry Gee

Turning down the lights

London

SPAIN has become the first country to take central government action to tackle the problem of the glare from artificial lighting that can wreck astronomers' observations. A new law introduced by the science minister, Javier Solana, will place strict controls on lighting on La Palma, Canary Islands, the site of several optical telescopes including the 4.2-metre Anglo-Dutch William Herschel Telescope.

Professor Alec Boksenberg, director of the Royal Greenwich Observatory (which runs the Herschel Telescope), says the new law is a real commitment, "not just empty words", as it sets up a fund of 300 million pesetas (£1.5 million) to pay for improvements to existing lighting over the next three years.

'Light pollution' has not yet created serious problems at La Palma, and the new legislation should avert any future difficulties. But US astronomers have already been badly affected. The Mount Wilson Observatory, once one of the world's leading astronomical facilities, can now usefully observe only the brightest stars because the lights of Los Angeles have brightened the night sky by a hundred times over the natural background.

US moves to fight light pollution have been taken at city and county level, through local ordinances controlling street lighting and illuminated advertising hoardings. David Crawford, of the Kitt Peak National Observatory in Arizona, says there have been some successes. During the 1980s, the sky above Kitt Peak

did not become significantly brighter, despite the growth of nearby cities.

But the skies above smaller observatories are still brightening rapidly, and Crawford has set up the International



The view from Kitt Peak National Observatory towards Tucson, Arizona. 'Seeing' was much better in 1959 (top) than in 1980. By the mid-1980s sky brightness had decreased as a result of controls on exterior lighting, although the town had continued to grow. (NOAO).

Dark Sky Association to publicize the problem. Energy-efficiency pressure groups are an important ally. The need is for street lights that direct light downwards, with little 'leakage' up into the night sky. The United States currently "spends more than a billion dollars a year just lighting up the sky", Crawford says.

Peter Aldhous

ARCHAEOLOGY

Investigation of Indian bones legal

Washington

ONLY a week ago, California archaeologist David Van Horne was facing a stiff prison term for sending two unidentified coin-sized bones to a laboratory to have them tested for human origin.

Tried under California law, which makes it a felony to desecrate Indian gravesites, Van Horne could have been sentenced to three years in a state penitentiary if found guilty. But in a precedent-setting decision, municipal judge B. J. Bjork late last week ruled that bones — even Indian bones — are "archaeological material" until they have been identified as human remains, and freed Van Horne.

The case had brought dozens of archaeologists — some with bumper stickers reading "Archaeology is Not a Crime" — to the courthouse in support of Van Horne, who has been a vocal critic of California's restrictive rules on investigating Indian remains. Van Horne was charged after his firm, Archaeological

Associates, had been hired to survey a construction site for possible Indian artefacts. He found two small bone chips that could not be positively identified as human, so he sent them to a University of California laboratory for analysis. When they turned out to be prehistoric human remains, Van Horne was charged with desecration of an Indian gravesite. The prosecution argued that he should have recognized the small charred bones as Indian remains.

Van Horne claims that he was picked on by local prosecutors for his views. "There is some question if this was malice on the part of the state", he says. In a previous case involving two small milling stones, he lost and was forced to rebury the stones, although he publicly ridiculed the state for spending \$150,000 to "bury two rocks". "It's a science versus superstition issue. It's nearly impossible to overcome the public guilt over native Americans", he says.

Christopher Anderson

Lower dose limits

London & Boston

THE International Commission on Radiological Protection (ICRP) has reduced its recommended limits for radiation doses received at work. Long-running studies have shown that the risks to health from radiation exposure are about three times higher than was thought in 1977, when the previous limits were issued.

The survivors of the Hiroshima and Nagasaki atomic bombs have been studied to estimate the risk of developing cancer through radiation exposure: as the population has aged, more survivors have succumbed to cancer than was expected. Recent animal studies have also shown that radiation is more likely to cause heritable genetic defects than was thought.

In 1977, ICRP recommended that workers should receive no more than 50 millisieverts (mSv) each year. This is still the maximum dose in any one year, but ICRP now also recommends that exposure should be kept below an average of 20 mSv per year over a given five-year period.

The ICRP limits have no legal standing, but they have provided an international standard on which national limits can be based.

The United Kingdom has been relatively quick to take note of its recommendations and adopted the 1977 limits as the legal limits in 1985. If the new recommendations are now adopted, they will affect only a small number of workers, most of them working either in the nuclear industry or in mines where radon gas might be inhaled. At the Sellafield nuclear waste reprocessing plant, for example, where radiation exposure has been linked to the occurrence of leukaemia in the children of workers in a controversial epidemiological study (see *Nature* **343**, 679; 22 February 1990), 44 workers received more than 20 mSv in 1989–90.

The United States has been much slower to take note of ICRP recommendations, and even now the Nuclear Regulatory Commission (NRC) is only just formally adjusting its radiation exposure limits for workers to correspond to the 1977 recommendations. Workers will be allowed no more than an effective dose equivalent of 50 mSv per year (or 5 rem).

The new limits will lower the maximum permissible annual dose of radiation, but critics in the United States complain that the new rules lag behind international standards and also that they allow loopholes that permit workers in some cases to be exposed to even larger doses of radiation than previously permitted by the NRC.

Peter Aldhous & Seth Shulman

Spending more to save . . .

London

SAVE British Science (SBS), the pressure group, this week demanded an increase of £1,300 million in government support for research and development, as well as far-reaching changes in British science education. In a document headed "Benchmarks for 2000", SBS also asks for 50 per cent increase in the numbers of science and engineering graduates, stipends for graduate students comparable with those obtainable at the beginning of "alternative careers" and a full-time Minister of Science and Technology.

SBS's argument is mostly based on comparisons between Britain and countries elsewhere, notably those of the European Communities (EC) and of the Organization for Economic Cooperation and Development (OECD). It says that the year 2000 has been chosen as a date by which its reforms could reasonably be implemented.

On education, SBS asks that all secondary-school students should be taught science by qualified teachers, and that secondary students should be encouraged to stay at school, with the intention that 80 per cent of those aged 16 to 18 should be in fulltime education.

Similarly, SBS asks that more than half of young people should be engaged in some form of higher education, not necessarily in traditional degree courses. The proportion of young people undertaking technical training should be increased by a factor of at least two or three, according to the document. It draws attention to the declining proportion of university teachers younger than 35, noting that recruitment for the future should begin now.

SBS also wants to see a 50 per cent increase in the numbers of graduate students who, it says, should be paid salaries appropriate to people beginning work as lawyers and accountants.

The case for increased support for civil research turns on comparisons with performance elsewhere, notably in Germany, Japan and the United States. Thus SBS wants spending on civil research and development to increase to 2.3 per cent of gross domestic product (GDP) by 1995, and to 2.7 per cent by the end of the decade.

The government's share of this (estimated to have fallen from 0.72 per cent in 1981 to 0.55 per cent now) would have to increase to 0.8 per cent in 1995, whence the £1,300 million a year. SBS also endorses the claims of universities and polytechnics for an injection of £1,000 million to replace obsolete equipment and buildings, and argues that restoring the capacity of the grant-making system requires an extra £400 million.

SBS says that industrial spending on civil research will have almost to double in numerical terms, implying by the end of the decade extra spending of £3,700 million a year on civil research and development funded by industry's own resources. It would be for the government to devise the appropriate incentives, SBS says.

That would be one task of the proposed Minister for Science and Technology. SBS says that it does not seek to be "over-prescriptive" at this stage about that person's role, but that such an appointment would be a "signal . . . that science policy is going to be taken seriously for the first time in Britain."

John Maddox

HIGH-ENERGY PHYSICS

HERA completed

Munich

HERA, the first electron-proton collider in the world, was completed at DESY (Deutsches Elektronen Synchrotron) in Hamburg earlier this month and is now being readied for start-up next year. The 6,336-metre-long oval tunnel is located 20 metres beneath the Bahrenfeld section of Hamburg.

The first particle beams at full energy are expected in the spring of 1991 and the first collisions in the summer. Protons will circulate at 820 GeV, electrons at 30 GeV, and collisions will be monitored by two large detectors known as ZEUS and H1.

HERA (Hadron Elektron Ring Anlage), which took six and a half years to build and cost approximately DM1,000 million (about \$675 million at current exchange rates), is the first European accelerator to use superconducting magnets.

The German Research Minister, Heinz Riesenhuber, last week ceremonially pushed a button to start the process of cooling the particles in the storage rings. Riesenhuber said that he was proud that HERA had been completed on time and within its budget.

HERA is expected to explore the structure of the nucleon in greater depth than ever before, and to investigate the weak force. Although current theoretical estimates indicate that the elusive top quark is a little too massive to be detected at HERA, researchers there hope that HERA's high resolution may allow them to determine if quarks are really the smallest constituents of matter.

Many countries contributed to HERA, both scientifically and financially, including France, Israel, Italy, Canada and the Netherlands. More than 700 researchers are working in high-energy physics and another 250 in synchrotron science at the facility.

Steven Dickman

Database for plant species

London

BOTANISTS are compiling a computer database of plant species names despite the chronic shortage of funds for taxonomic research. A meeting held on 12–13 November at the Royal Botanic Gardens, Kew, attracted delegates from 45 institutions to discuss Kew's own database initiative, the Species Plantarum Project (SPP). Grenville Lucas, keeper of the Herbarium at Kew, thinks that the SPP will cost some \$10–20 million, but the need for a taxonomic database is so great that botanists simply cannot wait for funds. "If we don't start now", Lucas says, "it'll never get done."

The first five-year phase of SPP will be a simple checklist of all 250,000 species of seed-plant, ferns and mosses. The second phase will be an in-depth treatment of basic plant taxonomy and biogeography, a taxonomic equivalent of the Human Genome Project. When complete, the SPP should supplant its eighteenth-century forbear, Linnaeus's *Species Plantarum*. SPP is just one of several ideas designed to cater for an increasing need for up-to-date taxonomic information. A meeting convened in Delphi, Greece last month (11–16 October) to discuss designs for a Global Plant Species Information System (GPSIS) was "hijacked", says Frank Bisby of the University of Southampton, by researchers who wanted immediate action. The result was the formation of the GPSIS Action Group, which is committed to producing a species checklist in just three years that might include fungi and algae as well as vascular plants and mosses.

Some of the work has already been done by researchers working on particular plant groups. Bisby's laboratory, for instance, is working on a checklist of the 16,500 species of legumes (one-twelfth of all species of flowering plant) as a part of the International Legume Database and Information Service (ILDIS). Bisby views the ILDIS checklist as a good working model of the GPSIS initiative, but is angry about the dearth of secure funding for taxonomy: seen, he says, as of "marginal interest to many people, but a lot of interest to nobody".

Bisby thinks that the SPP and the GPSIS Action Group will join forces, given that their aims are virtually identical. But once a database is up and running, funds will be needed to maintain and update it. "The management of it is going to be a monster" concedes Lucas, who is nevertheless confident of success with "the fraternity of taxonomy worldwide shoving behind the wheel".

Henry Gee

Sexual lifestyles under scrutiny

K. Wellings, J. Field, J. Wadsworth, A.M. Johnson, R.M. Anderson and S.A. Bradshaw

Much useful information can be obtained from a knowledge of people's sexual habits, not least information about AIDS. Such studies should be judged by scientists, not by politicians.

"HUMAN sexual behavior represents one of the least explored segments of biology, psychology and sociology. Scientifically, more has been known about the sexual behavior of some of the farm and laboratory animals." Alfred Kinsey's words¹ remain as true today as they were 40 years ago. Despite an apparent increase in openness about sexual matters, there are few reliable data on the sexual habits of the general population. A cursory glance at the literature seems to suggest that there is an abundance of material, but most of this is qualitative, rather than quantitative, and the field is one in which modern methods of population survey research have seldom been applied. No one has progressed from the pioneering work of Kinsey and colleagues in the United States, which was in any case based on volunteer samples and therefore may not be representative of the general population.

The lack of reliable information has serious implications for understanding fertility patterns, trends in use of contraceptives and the epidemiology of sex-related diseases (cervical cancer, for example). But the emergence of AIDS, whose causative agent, the human immunodeficiency virus (HIV), can be transmitted by sexual contact, provides a new sense of urgency, not only among epidemiologists and those in health education, but also for the development of methods for quantification of behavioural patterns².

Hostility

Why has the study of sexual behaviour attracted little funding or serious attention? Researchers have been understandably chary of entering a field more likely to cast a slur than to confer status on their professional reputation. The history of sex research is replete with examples of hindrance, obstruction and discouragement. Questionnaires have been seized as pornographic, researchers slanderously attacked and their findings suppressed lest they threaten the established moral order. Kinsey, for example, met with hostile reaction from colleagues and threats of dismissal¹; and Lanval, a Belgian researcher working in the 1930s, worked at night to avoid police raids³.

The situation is little changed today. The advent of AIDS has provided some impetus and legitimized the study of sexual behaviour — a research protocol for

surveying sexual lifestyles has been developed by the World Health Organization and surveys are already under way in several European countries. Yet researchers on both sides of the Atlantic have encountered difficulties in obtaining funds to mount large-scale cross-sectional surveys of sexual behaviour. Fieldwork for the main stage of the British survey of sexual attitudes and lifestyles, set to begin in April 1989, was delayed because of a much-publicized government veto on financial support. And those planning the US study still await a decision as to whether they can continue.

The British survey was conceived in 1987 with the aim of providing data to assist in the prediction and prevention of future spread of sexually transmitted diseases, in particular AIDS and HIV-related diseases, using a random sample of 20,000 of the general population. A two-year development programme including qualitative work to aid questionnaire design, extensive piloting of the research instrument and a large-scale feasibility survey based on a random sample of 1,000, culminated in readiness for the main stage of the survey. This was the point at which further work was halted as a result of political pressure. Nevertheless, following scientific review, the Wellcome Trust, a British charity which finances medical research, provided money for the work to continue.

The objections to the survey were reportedly threefold. It was said that such an enquiry represented an invasion of people's privacy; that it would be unlikely to produce valid results; and, as many AIDS-related surveys of sexual behaviour were already under way, that it would be an inappropriate use of public money. The decision provoked the dismay, not only of the researchers involved, but also of other scientists.

It is difficult to understand qualms over spending public money on general population research when numerous studies of sexual behaviour of minorities have been funded without any objections to intruding into the private lives of homosexual or bisexual men, prostitutes, drug users or teenagers. The AIDS epidemic has undeniably led to a huge proliferation of research activity in the area of sexual behaviour — enough to prompt one journalist to describe it as an "overcrowded field". Yet so far, studies have focused largely on specific subgroups of

the population, identified because they occupy a unique or crucial position relevant to understanding the aetiology of the disease. However valuable their findings, studies of this kind are hampered by the absence of base rates against which to assess the meaning of the data collected and the extent to which they can be generalized to the wider population. A cross-sectional survey is vital to provide the population parameters required to set other studies in context.

Acceptability

Concern for the acceptability of the survey and the validity of its findings is entirely legitimate, and indeed was uppermost in our minds. Sexual behaviour is regarded as intensely personal by most people: reported accounts of behaviour have to be relied on, there are few possibilities for objective verification, and disclosure may include reports of acts which are socially disapproved of, if not illegal. The twin issues of veracity and recall particularly exercise those who have doubts about the validity of sexual behaviour surveys. Yet problems relating to people's ability to recollect accurately and report honestly beset investigation into many aspects of human behaviour, including drinking and smoking habits, disclosure of earnings, views on racial minorities and so on. Further, such questions are more properly the concern of science than of politics. The question of whether a survey can be designed to produce reliable data is one which can be answered only after careful and thorough preparatory work and rigorous analysis of a feasibility study. We were aware of the problems of validity, so our long period of development work concentrated largely on attempts to minimize their effect. We describe some elements of this process below.

A number of features can be built into the methodology to increase the likelihood of acceptance. A non-judgemental approach on the part of the interviewer and assurance of confidentiality are essential, as is an understanding on the part of the respondent of the urgent need for the data and the credentials and integrity of the originators. In the case of the British survey, data collection was reportedly made easier by reference to the need for the information in the public-health context of the AIDS epidemic and the fact that the survey was publicly funded. Most interviewers found the feasibility survey

no more difficult to carry out than other social surveys. Female interviewers generally achieved higher response rates than males, but respondents claimed a non-judgemental approach to be more important than interviewer characteristics such as gender and age.

Qualitative development work undertaken to assist questionnaire design revealed some discomfort, on the part of both interviewer and respondent, at face-to-face disclosure of intimate information. We took account of these sensitivities by including questions of a less personal nature — on general health, family circumstances, sex education, early sexual experience and so on — in the first part of the interview. More sensitive questions on, for example, numbers of sexual partners, frequency and nature of different sexual practices, history of contact with prostitutes and drug injection were answered in a self-completion booklet, unseen by the interviewer, sealed by the respondent and identified only by a number. Those who had had no sexual experience were not asked to complete the booklet. The confidentiality achieved by this procedure emerged as an important reason for compliance. The use of a combined approach also provided an opportunity to compare responses on items repeated in both self-completion and face-to-face interviews (for example, homosexual experience), and the choice of format was ratified by a higher proportion of respondents reporting same-sex experience in self-completion than face-to-face interview.

Of crucial importance is the way in which questions are phrased and posed. Discomfort with, and misunderstanding of, language will jeopardize willingness to respond and the ability to produce accurate responses. Many researchers, including Kinsey, have cautioned against the use of a standardized questionnaire and neutral terminology, advising instead the use of the vernacular. But we found that the use of the respondent's own language was unlikely to provide either the appropriate rapport with the interviewer or the required standardization. Sexual behaviour is rarely spoken about publicly so the language used to describe it is often inadequate and inappropriate. Many terms used in the vernacular in the English language double as terms of abuse, and their use in the research setting could cause offence. Most respondents expressed a preference for the use of formal terms (sexual intercourse, penis, vagina). We also found there were misunderstandings, among a sizeable minority, of several terms widely used in the health-educational literature, if not in everyday parlance — penetration, vaginal and heterosexual sex — sufficient to threaten substantially the overall validity of the response. There was also a tendency to

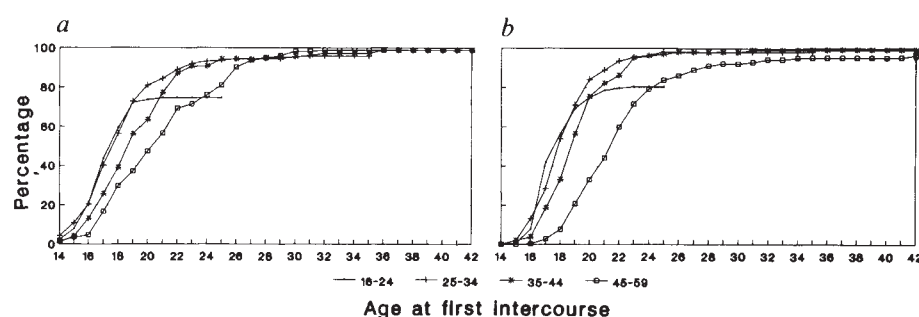


FIG. 1 Median age at first heterosexual intercourse reported by *a*, men, and *b*, women, in four different age groups. (Results are from the feasibility study described here.)

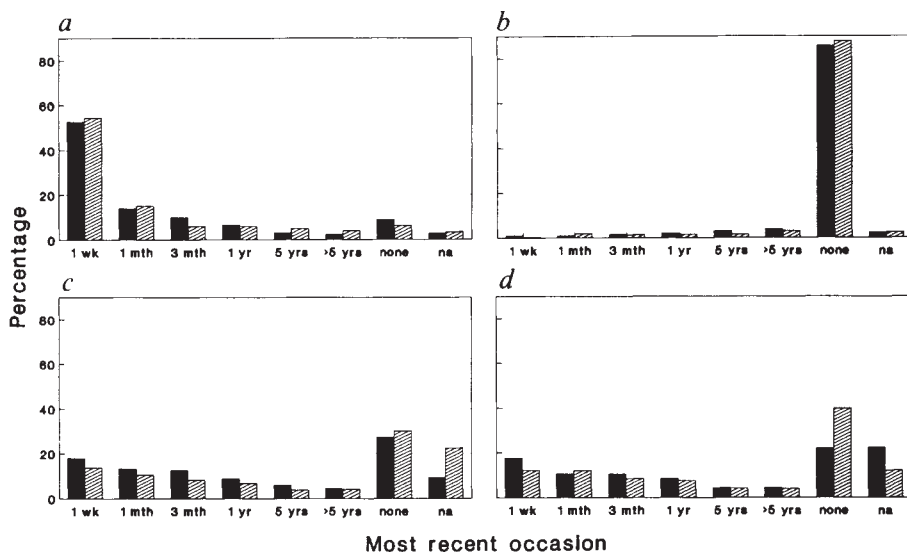


FIG. 2 Reported incidence of different types of sexual intercourse. *a*, Vaginal intercourse reported by men and women. *b*, Anal intercourse reported by men and women. *c*, Oral intercourse male to female. *d*, Oral intercourse female to male. Solid bars are reports by men, hatched bars those by women; na = not answered. (Results are from the feasibility study described here.)

attribute to unfamiliar terms meanings associated with unorthodox or bizarre sexual practices.

In addition, there was wide variation in the meaning attached to crucial variables such as 'sexual partner' (some respondents, for example, discounted their spouse as a 'sexual partner') and 'having sex' (sex was equated with 'intercourse' for heterosexuals, whereas homosexual respondents included a broader repertoire of sexual acts in the definition). Other terms such as 'masturbation' were simply disliked. As a result, we defined words carefully and precisely, not simply in terms of a dictionary definition, but in terms of a specified practical reality. A glossary was included to give a standard interpretation. 'Sexual partners', for example, we defined as "people who have sex together just once, or a few times, as regular partners or as married partners".

The literature on memory research confirms that incidence is easier to recall than frequency, so that events occurring once tend to be most memorable. Events such as first intercourse seem sufficiently salient to be easily called to mind; fewer than 2% of respondents in the feasibility study failed to remember their age at first inter-

course. More difficult to remember are events which accumulate over a long period of time, such as numbers of sexual partners, particularly for individuals with more varied sexual histories — yet this is a crucial variable.

As forgetting may be partly a process of blocking, some of the methodological devices designed to deal with the problem of honesty — the formulation of neutral questions, a permissive interviewing style, and provision of assurances of confidentiality and anonymity — can help the respondent to retrieve information. Other design features can facilitate the reckoning process. Scheduling the order of questions so that those about events such as first sexual experience and childbirth are asked first provide a framework in which respondents can locate less easily recalled experiences, and seems to trigger associations. The sequence in which different time periods are set out in the question format also influences recall, and the feasibility survey provided the opportunity to perform split-run experiments exploring the extent of order effect.

A significant shortcoming of previous surveys of sexual behaviour has been their failure to use probability sampling pro-

cedures. Kinsey's studies were based on volunteer samples, whereas studies of homosexual men have generally relied on self-identified samples, from gay clubs, for example. Market-research agencies commonly use quota sampling (generally using age, sex and social class to construct a representative population sample). Volunteers are more likely to self-select according to the variable under investigation (sexual behaviour), and so cannot be taken to be representative.

Feasibility

The design of the feasibility survey was a multi-stage random sample stratified by socio-economic characteristics of the neighbourhood using the Post Office's small-user postcode address file for the United Kingdom. Each address was approached by an interviewer and all members of the household aged 16–59 enumerated, one individual being chosen from the list by a random selection procedure and invited for interview. The achieved response rate of 65% is similar to that of other random sample surveys on sexual behaviour in Western countries, and of previous surveys in the United Kingdom, which range from 60 to 68% (refs 4–8). Higher response rates have been achieved in studies in which limited questions on sexual behaviour formed a small part of the questionnaire^{9,11}, and in some of these a 5% drop in item response rate is reported at the point at which such questions are introduced.

The extent to which the sample is representative of the general population is important to the reliability of any survey such as this. Representativeness in terms of demographic variables such as age, social class, marital status and ethnic origin is easily assessed using available census data. Compared with the UK 1986 population projection based on the 1981 census and the 1985 General Household Survey, our sample was broadly representative of the demographic structure of the population with the exception of older men aged 45–59. A slightly higher refusal rate on the part of older men is common to studies from other countries and confirms our decision to restrict the survey to the age group 16–59.

A further assessment of the representativeness of data is achieved by comparing responses with independent data sets, such as comparisons of self reports of sexually transmitted diseases with national statistics. For example, in our study, 1.4% of respondents reported visiting a sexually transmitted disease clinic in the past year and 3.4% in the past 5 years. For England and Wales, this provides a population estimate of 403,000 attenders per year. This is consistent with an estimate of 332,000 per year based on a study carried out in 1978 (ref. 12).

Clearly, the absence of reliable base-

rate data makes it more difficult to determine whether the sample is representative in terms of sexual characteristics. Nevertheless, to take the variable of sexual orientation — again, of central importance in the context of this survey — the proportion of men reporting a same-sex experience and at least one male partner is much higher than in earlier pilot studies and is consistent with recent studies in other Western countries^{6,10}. In our sample, 9% of men and 4% of women reported some homosexual experience, and 5% of men and 1% of women reported ever having a homosexual partner. Reports of the prevalence of socially stigmatized behaviours, such as homosexuality and anal intercourse, must be regarded as minimum estimates. Nevertheless, the higher rates of homosexual experience compared with earlier pilot work (where only 0.5% of men reported any such experience¹³) reflect the greater subtlety and clearer definition in question wording in this study, and emphasize the importance of question design and delivery in eliciting responses to sensitive questions.

Analysis of the results of the feasibility study was limited to attempting to demonstrate the extent of validity and reliability. Nonetheless, even this preliminary analysis shows that the study has so far successfully captured the marked generational changes in sexual attitudes and behaviour. For the youngest age group of women (16–24-year-olds) in the study, median reported age at first intercourse, for example, was 4 years earlier than it was for the oldest group (45–59-year-olds) (Fig. 1). Of particular interest to us in attempting to assess veracity of response were activities surrounded by strong taboos and therefore most likely to be under-reported. Anal sex between a man and a woman is socially censured and so is a useful indicator of our capacity to elicit honest responses: respondents showed no discomfort at being asked the question and a sizeable minority reported the experience (Fig. 2).

Internal consistency checks are useful in checking veracity and recall. We were encouraged by the consistency with which respondents completed questions, as well as the consistency between men and women in reporting sexual acts. In a representative sample of heterosexuals, reports of men and women on the frequency of specific acts over short times should concur, and the most recent occurrence of vaginal and anal intercourse were almost identically reported by men and women. Reporting of oral sex was less consistent, but an ambiguity in the question could have been responsible.

Men had a tendency to report more partners than women, a finding which this study shares with the US Household Survey and others^{6,10}. Explanations for this have been discussed elsewhere¹⁴ and

include considerations of patterns of sexual and age mixing within and outside the population studied. In our data, the difference between men and women is small for finite intervals of 5 years or less but greater over lifetimes. The influence of social gender expectations on estimates of numbers of sexual partners is undeniable but it seems to operate chiefly at the point at which memory decay starts to influence the accuracy of estimates. Nevertheless, our data are comparable with data from the US Household Survey in showing striking similarities in numbers of partners reported by men in the past year.

Reliability

As a result of the development work, we believe we have been successful in gaining the cooperation of the public to contribute reliable reports of sexual behaviour, to justify proceeding to the main stage of the survey. Ideally, this initial attempt will be the first stage of a continuous process of data collection. Whether the controversy surrounding the survey has had adverse consequences for the main study remains to be seen. It would certainly be regrettable if publicity surrounding the political decision should reinforce just those taboos which hindered attempts at scientific research in this area in the first place □

K. Wellings and J. Wadsworth are in the Department of Public Health, St Mary's Hospital Medical School, Norfolk Place, London W2 1PG, UK. J. Field is at Social and Community Planning Research, 35 Northampton Square, London EC1V 0AX, UK. A. M. Johnson and S. A. Bradshaw are in the Department of Genitourinary Medicine, University College and Middlesex Hospital School of Medicine, London W1N 8AA, UK. R. M. Anderson is in the Department of Biology, Imperial College, London SW7 2AZ, UK.

1. Kinsey, A.C., Pomeroy, W.B. & Martin, C.E. *Sexual Behavior in the Human Male* (Saunders, Philadelphia, 1948).
2. Anderson, R.M. & May, R.M. *Nature* **333**, 514–519 (1988).
3. Lanval, M. *An Inquiry into the Intimate Lives of Women* (Cadillac, New York, 1950).
4. Gorer, G. *Sex and Marriage in England Today* (Nelson, London, 1971).
5. Schofield, M. *The Sexual Behaviour of Young People* (Longman, London, 1965).
6. Sundet, J.M., Kvale, I.L., Magnus, P. & Bakkeley, L.S. in *The Global Impact of AIDS* (eds Fleming, A.F. et al.) 53–60 (Liss, London, 1988).
7. Ross, M.W. *J. Sex Res.* **25**, 441–450 (1988).
8. Gotthberg, M., Giesecke, J., Scalia Tomba, G. & Tull, P. *MME Rapport No 12* (Stockholm, Sweden, 1989).
9. Turner, C.F., Miller, H.G. & Moses, E.M. (eds) *AIDS, Sexual Behaviour and Drug Use Ch. 2* (National Academy Press, Washington, DC, 1989).
10. *Mortal. Morbid. Weekly Rep.* **37**, 565–569 (1988).
11. Forman, D. & Chilvers, C. *Br. med. J.* **298**, 1137–42 (1989).
12. Belsey, E.M. & Adler, M.W. *Br. J. ven. Dis.* **57**, 285–9 (1978).
13. Johnson, A.M. et al. *AIDS* **3**, 135–141 (1989).
14. Johnson, A.M. et al. *Nature* **343**, 109 (1990).

ACKNOWLEDGEMENTS. We thank the Economic and Social Research Council and the Health Education Authority for funding the development work and the Wellcome Trust for funding the main study. We also thank Roger Jowell, Michael Adler, David Miller and Paul Elliott for advice, and our interviewers and study participants.

Test ban to go comprehensive?

SIR—The differing perceptions of Mexico and the United States prevented the production of a final document at the Nuclear Non-Proliferation Treaty (NPT) Review Conference in Geneva from 20 August to 14 September. But was Mexico's protest really "unrealistic" and "counterproductive"? ("NPT in serious trouble", *Nature* **347**, 213; 1990)? Has it harmed the treaty to the extent that the principle of consensus needs to be broken, for example by making use of the amendment process? Has this meeting set a precedent for failure of the review and extension conferences of the treaty in 1995?

A crucial issue is whether pressures to amend the NPT will strengthen or destroy the treaty. It was agreed by all participants in the 1975 review conference (which itself nearly ended in deadlock) that every effort should be made to reach decisions by consensus. It was thought that any decision arrived at by simple majority (even near-unanimity) would be an imposition on those losing the vote and would lead to states leaving the NPT as permitted under article X. It would thus be a greater indication of failure than non-agreement on a final document.

On listening to the speeches in the plenum of the review conference, one fact becomes plain: everybody thinks the NPT is a valuable and important treaty that should be preserved. Everybody speaks of the necessity of its extension.

The question is: what measures are required to ensure extension for a meaningful period beyond 1995? The non-aligned states, including Mexico, answered that it was the conclusion of a Comprehensive Test Ban Treaty before 1995.

Since 1981, neither the United States, the Soviet Union nor the United Kingdom has participated in such negotiations. Calls by more than a third of the parties to the Partial Test Ban Treaty to amend it to a comprehensive treaty or secure a mandate for the Conference on Disarmament *ad hoc* committee to negotiate a test ban show the degree of frustration felt by many states.

The United States will go to the amendment conference in New York on 7–18 January facing difficult if not insuperable diplomatic problems. Rumours are circulating that the United States may walk out after the first day. For the non-aligned, the problems of ensuring unity are just as important. At least one ambassador to the review conference from a prominent non-aligned country has expressed his anger at the Mexican delegation's stance.

A Partial Test Ban Treaty amended to a comprehensive treaty would halt the qualitative improvement of nuclear

weapons and would do much to safeguard the future of the Nuclear Non-Proliferation Treaty.

DAVID LOWRY
PAUL HELLIWELL

*European Proliferation Information
Centre,
258 Pentonville Road,
London N1 9JY, UK*

Problems of UK universities

SIR—There is no doubt that British universities could and should do more to recover a greater fraction of overheads from external research contracts than the 14 per cent reported by Science and Engineering Policy Studies Unit (SEPSU) (*Nature* **347**, 506; 1990). Something like the old University Grants Committee estimate of 40 per cent would certainly make our university finance officers happier, but according to SEPSU even this is far too low. In strict accountancy terms, I do not doubt that SEPSU can justify this, but how do they allow for the hidden benefit (overhead) that accompanies most external funding to universities?

Many 'soft-money' university researchers usually find time within their contract to pursue aspects of related, non-contract research which enhances their academic progress, and that of their department and discipline. Furthermore, many take an active role in undergraduate and postgraduate education and training. Without their input, the staff funded by the Universities Funding Council (UFC) would find it even more difficult to cope with their academic duties in the face of declining support from government.

It would seem that any estimate of research overheads should allow for this, by deducting what it would cost the universities to provide UFC-funded staff to take over the voluntary academic duties of our soft-money colleagues.

MICHAEL J. STOCK

*Department of Physiology,
St George's Hospital Medical School,
London SW17 0RE, UK*

SIR—Your leading article on undergraduate university funding in the United Kingdom (*Nature* **345**, 462; 1990) and the current changes proposed in the school syllabus prompts me to raise two fundamental issues.

First, will the accepted decline of the factual scientific knowledge of the school-leaver affect the length, content and standard of degree courses? The answer must be yes, particularly where the acquisition of a basic 'data base' is paramount, for example basic medical science

and the natural sciences. In turn, this can only increase the length of PhD registration.

The second issue is more sinister. It has been suggested that the measure of success in school can be measured from the percentage of higher grades awarded. It does not require a quantum leap of the imagination to see that university funding may also depend on such perceived success — the number of firsts awarded. The result may well be an erosion of standards that will adversely affect the academic competitiveness of the United Kingdom.

ANJAN BANERJEE

*Department of Clinical Biochemistry,
King's College Hospital,
Denmark Hill, London SE5 9PJ, UK*

The testing of laetrile

SIR—John Cairns (*Nature* **345**, 585; 1990), in reviewing Ralph Moss's book *The Cancer Industry*, says that Moss thought that the Sloan Kettering Center (SKC) "trials of laetrile had been rigged to give a negative result. This explains the theme in most of the chapters which is that . . . probably some wondrous cure for cancer . . . is being suppressed by the chemical industry and the cancer establishment After all if something like laetrile really worked . . . it is hardly conceivable that the many people who have been given laetrile would keep silent"

Laetrile is a poor example for Moss's thesis. Laetrile (amygdalin) is a cyanogenic glycoside that was proposed for cancer treatment on the basis that it would be broken down by an enzyme in cancerous tissue to liberate cyanide which would "kill the cancer". This hypothesis was falsified by D. M. Greenberg's experiments¹, after which the proponents of laetrile alleged, without any evidence, that it was "vitamin B₁₇", a deficiency of which caused cancer². This fantasy was used to promote laetrile for cancer, and political pressure was exerted on the National Cancer Institute to schedule tests at four prominent cancer centres including SKC and Mayo Clinic. Not one patient was cured or even stabilized. The trials were not "rigged". Indeed, it is hard to conceive of a test of laetrile being "rigged", except by omitting the medication. The "many people who have been given laetrile" did not all "keep silent". Some of them or their surviving relatives have participated in various court cases in which punitive action was brought against the dispensers of laetrile³.

THOMAS H. JUKES

*University of California,
Berkeley, California 94720, USA*

1. Greenberg, D. M. *Western J. Med.* **122**, 345–348 (1975).
2. Jukes, T. H. *Nature* **269**, 283 (1977).
3. Yarbrow, J. W. *Missouri Medicine* **76**, 195–206 (1979).

Why become complicated?

SIR—P. R. Sheldon (*Nature* 347, 704; 25 October 1990) asks why simple organisms such as bacteria do not develop the same complexity as other less numerous lineages, even though their prolific numbers should allow more variations to develop. The simple answer is that eukaryotic organisms have more DNA than prokaryotes and so more potential for development, although this does not explain why prokaryotic DNA is restricted.

One possibility is that the mitochondrion, by providing an organelle for the efficient production of energy-related proteins, allowed the eukaryotic cell to develop more complex control mechanisms — and hence a more complex genome — than a prokaryotic cell. This would also explain how a chance combination of two prokaryotic cells could be selected as a useful symbiont (L. Margolis, *Symbiosis in Cell Evolution*, Freeman, 1981).

In prokaryotes, the energy-producing proteins are manufactured alongside less-used proteins and are subject to the same manufacturing constraints. In eukaryotes, the mitochondrion contains only a little DNA which it uses to manufacture proteins using ribosomes and a reduced number of transfer RNAs. Hence the overheads for protein production are a minimum and the speed of protein production should be a maximum. The main eukaryotic cell is then freed from making these proteins and has the opportunity to specialize in making a wider range of proteins in a slower, more controlled, fashion.

Which protein manufacturing system most closely resembled that used by the early prokaryotes: the cell or the mitochondrion?

PHILIP WOOD

28 Park Lane,
Wembley HA9 7RZ, UK

SIR—The answer to Sheldon's question, I believe, is that only eukaryotic cells can evolve complex morphology. It is well known that all prokaryotic organisms are unicellular and morphologically simple, whereas even unicellular eukaryotes are morphologically more complicated. One of the reasons for this difference is mitosis and meiosis (which occur in eukaryotes but not in prokaryotes), and another is the presence of endoplasmic reticulum in eukaryotes, on which are enzymes needed for morphogenesis and multicellular organization.

But the shortfall in this explanation is that it fails to deal with the question of what are the needed enzymes and how do they work — what, in other words, is the biochemical basis of morphogenesis? We

can expand this question: what is the biochemical basis of cellular irritability in all its manifestations? This is a fundamental question, for which present-day molecular biologists have no answer.

BENJAMIN LOWENHAUPT

Biology Department,
Allegheny College,
Meadeville, Pennsylvania 16335, USA

Cohort of genes

SIR—I was pleased that *Nature* (347, 221; 1990) highlighted the fact that confidentiality, in particular matters of life and health insurance, will complicate attempts to sequence the human genome in ways that should be faced. What the report did not address was a much more fundamental and equally urgent issue of how we are going to discover what effect specific common genetic variations have on health and development. The report simply states that "with good planning and much hard work there will be a comprehensive listing of all functioning human genes and information about the ways in which they can vary either harmlessly or otherwise, from one individual to another". There is, however, a conspicuous lack of forward planning of research that will have the power and precision to determine whether a given genetic variation is harmless or otherwise, and, if otherwise, in what circumstances it is harmful. The author assumes (probably correctly) that the overwhelming proportion of genetic variations will be harmless, and no doubt also assumes that clinical research will easily sort out those that are harmful. In fact, this will not be easy, particularly with respect to which DNA sequence polymorphisms confer susceptibility or resistance to common multifactorial disease.

The growing tide of anxiety about being singled out for genetic tests, in the absence of clear clinical indication, will, I predict, make the analysis of controls in conventional case control studies problematic, especially in childhood where additional ethical constraints operate. Furthermore, case control design, although often cost-effective, is unsuitable for disclosing positively protective alleles (not just the absence of risk alleles); just the type of genetic variation for which there would be positive selection pressure.

What is needed is a population cohort study in which a cell line on each individual provides a renewable source of DNA for analysis and where it is accepted by everyone involved that unspecified and strictly confidential research on factors influencing health and development will be performed. Such a study, the Avon Longitudinal Study of Pregnancy and Childhood has just been launched, but, as yet, the blood collection next spring and the generation of 22,000 cell lines from cord and maternal bloods

are without funding. This unique opportunity to make sense of the genetic variation revealed by the Human Genome Project will be virtually impossible to recreate.

What will be the use of a comprehensive listing of DNA sequence polymorphisms within all functioning human genes, if we cannot discover what they mean in terms of health and development? It will be impossible to sustain support for the Human Genome Project if this aspect is not addressed soon and comprehensively.

MARCUS E. PEMBREY

Mothercare Department of
Paediatric Genetics,
Institute of Child Health,
University of London,
London WC1N 1EH, UK

Art for art's sake

SIR—Your admiring caption to a colour picture of a fine feather tabard from ancient Peru (*Nature* 347, 425; 1990), currently for sale by a London art dealer, noticed that "primitive" art is gaining in status and value. Thomas Gibson Fine Art, vendors of the tabard, assure me that their items were legally exported from Peru and that they can supply a purchaser with a written statement to that effect. But the market in antiquities as art objects inflicts devastation on the archaeology of many countries, notably Peru. Site after site has been quarried by clandestine diggers, wrecking all archaeological context and stratigraphy in quest of saleable 'goodies', which are smuggled abroad and leave the possibility of archaeological understanding in ruin. European looting of the heritage of native Americans goes back centuries, but the commercial trade is larger and the effect more dreadful now than ever (T. Kaiser *J. Field Arch.* 17, 205–10; 1990). Every pillaged site in the Peruvian desert is a loss to science and to the people of Peru whose history it held. The same goes for those many other places — from the Greek Cycladic islands to the desert southwestern United States — whose ancient artefacts chance to catch the modern fancy.

Many of the antiquities that circulate in the art trade are looted; and the promotion of their elegant value is the engine of price that fuels the looters and smugglers — as surely as a good market in the developed countries for tropical hardwood finances the cutting of the rain forests.

You call the tabard 'primitive' art, and the appropriate *Oxford English Dictionary* definition of primitive is "simple, rude, or rough". The tabard is not primitive. It simply comes from an artistic tradition outside the Western experience.

CHRISTOPHER CHIPPINDALE

Antiquity,
85 Hills Road,
Cambridge CB2 1PG, UK

The new biology of immune recognition

A deeper understanding of antigen-specific immune responses is fast emerging. Last week, at *Nature's* conference "New Horizons in Immunology", there were reports of progress in unveiling the fundamental mechanisms, and of the latest attempts both to increase the versatility of responses and defeat them in the battle against autoimmunity.

Boston

MUCH of the research that has stripped the mystery from antigen-specific recognition by T lymphocytes reflects the impact of molecular technology, with the help of which it has been possible to probe the development of tolerance in T-cell transgenic mice and to make soluble human major histocompatibility complex (MHC) proteins in sufficient quantity for X-ray crystallography. Opening the conference, Jack Strominger (Harvard University) pointed out that computer technology too has played a part — not only in helping to solve the crystal structure of HLA-A1, but also, by means of computer graphics, through removing the antigenic peptide from the cleft in the molecule in which it binds (a feat, as he remarked, that is hard to achieve experimentally). This, as emphasized by the two speakers who followed him, reflects the singular status of the bound antigenic peptide as effectively an intrinsic part of the folded structure of the MHC molecule. The formation of the complexes between class I and class II MHC molecules and antigenic peptides that are recognized by T cells was the focus of the first session.

Peptides and pockets

The function of class I MHC molecules is to pick up peptides derived from the proteins of intracellular parasites, such as viruses, and display them on the cell surface for recognition by cytotoxic lymphocytes capable of destroying the infected cell. Exactly how and where the MHC-peptide complex forms has however been unclear, because the viral proteins are presumably made in the cytoplasm whereas the MHC molecule is synthesized on the endoplasmic reticulum with its peptide-binding domains in the lumen.

A possible answer to the problem is to be found in a mutant mouse cell line, RMA-S, in which the two chains of the MHC class I molecule fail to associate and fold correctly. Alain Townsend (University of Oxford) showed that class I molecules in this cell line can be rescued if they are 'fed' an appropriate antigenic peptide — thereby providing the first evidence for the crucial role of the peptide in their folding — and suggested that the mutation in RMA-S might be in a protein that transports peptides into the lumen of the endoplasmic reticulum. He was able to report in Boston that John Trowsdale, exploring the MHC class II region in man

to which the mutation maps in mouse, has identified a gene encoding one of the so-called ATP-binding cassette family of membrane transporter proteins to which the multi-drug resistance and cystic fibrosis proteins belong; and that although at this stage the evidence is only circumstantial, it seems at least possible that this is the transporter protein. The same gene has been identified in various species by at least three other groups, and complementation tests with transfected cells should soon establish whether it is the cause of the RMA-S defect.

A second mysterious feature of the binding of peptide by MHC molecules is its promiscuity, in apparent violation of the lock-and-key principle of receptor-ligand interactions. Don Wiley (Harvard University) however pointed out that the mystery largely evaporates if the peptide is seen as an integral component of the folded protein, there being many instances of secondary-structural components of proteins with little or no sequence similarity. Nonetheless, class I molecules do bind peptides selectively, a property of vital importance to the hope of designing anti-viral peptide vaccines, and co-crystals of MHC molecules with identified peptides should help to establish the basis for selective binding. It is therefore an important achievement of Wiley's laboratory to have found a way of reconstituting complexes of MHC class I molecules with specific peptides from purified components.

In the absence of a crystal structure, Wiley did not speculate as to whether class II MHC molecules, which have a different chain composition from class I but homologous domain structure, have an analogous relationship with their bound peptides; in the meantime, most of the controversy on class II seems to centre on the issue of whereabouts in the intracellular vesicular system they lose the 'invariant chain' that protects their binding pocket from occupation by endogenous peptides, and collect the appropriate fragments of endocytosed exogenous proteins. Peter Cresswell (Duke University) recapitulated the evidence from his laboratory and Frances Brodsky's that association with antigen occurs in an early endosomal compartment, whereas Emil Unanue reported new evidence indicating that some proteins may progress as far as the lysosomal compartment for degradation before associating with class II. These experiments,

which involve different cell lines and different proteins, are not mutually exclusive — given the history of immunological controversies, it is as well to bear in mind the possibility that both results are right.

Regulating the repertoire

The receptor for antigen on T lymphocytes has evolved to recognize the complex of MHC and antigenic peptide. But the receptor repertoire must be selected during ontogeny to ensure that self MHC molecules are recognized when their antigen pockets are occupied by non-self peptides but not when they are occupied by self peptides. The process is achieved most reliably by deletion of potentially self-reactive cells during thymic ontogeny, but it cannot be expected to operate for self peptides that are not expressed in the thymus. One mechanism for dealing with this (Ronald Schwartz, NIH) may be so-called clonal energy — the inactivation of cells that encounter antigen in the absence of a second signal normally required for a response to an infectious agent. This state has been induced in T cells *in vitro* in Schwartz's laboratory, but can be reversed after proliferation stimulated by interleukin-2, which he speculates may serve to dilute a repressor. If analogous processes can indeed occur *in vivo*, they may explain how tolerance can be broken and autoimmune disease ensue.

A second mechanism of autoimmune disease is suggested by observations in Philippa Marrack's laboratory (National Jewish Hospital, Denver) on the usage of a specific T-cell receptor β chain in the T cells associated with the inflammatory reaction in the joints of patients with rheumatoid arthritis. Marrack finds a very high proportion of V β 14 compared with that in peripheral blood cells. The biased V β usage suggests to her the operation of a 'superantigen' — one of a group of molecules characterized by their ability to bind specific β chains of the T-cell receptor independently of either the α chain or the peptide occupying the pocket. A superantigen encoded by an infectious agent, as some are, could stimulate potentially self-reactive cells able to recognize tissue-specific determinants in the joint.

John Kappler (National Jewish Hospital, Denver) raised the possibility that the endogenous superantigens of mice, which were the first to be discovered but have remained unidentified except genetically, are encoded by endogenous retroviruses.

In strains of mice carrying the mouse mammary tumour virus, but not in others, there is thymic deletion of cells bearing a specific V β chain, a phenomenon generally diagnostic of the need for tolerance to a superantigen. If indeed it is the viral sequences that encode the superantigen then it may be possible to identify an endogenous superantigen of mice.

If all potentially self-reactive T cells are deleted, inactivated or suppressed, and given that most or all B-cell responses depend on T-cell help, are direct mechanisms of B-cell tolerance necessary? Polly Matzinger (NIH) argued that they are, because the determinants recognized by B cells are quite different from those recognized by T cells, and it is therefore imaginable that a T cell could provide help to a B cell in producing antibodies against a self determinant it did not recognize. Moreover, antibodies, unlike T-cell receptors, undergo somatic mutation so that the emergence of self reactivity is a risk throughout life.

Christopher Goodnow (Stanford University) directly addressed the question, describing the B-cell anergy induced in transgenic mice that express hen egg lysozyme in all their cells and the antibody recognizing it on all their B cells. Like the T-cell anergy described by Schwartz *in vitro*, B-cell anergy can be broken, either by 'parking' the anergic B cells in mice not expressing hen egg lysozyme (although reactivity is not fully restored in these conditions) or, as in Schwartz's experiments, by stimulating proliferation with a mitogen (in this case LPS) *in vitro*.

Defeating diversity

Given that tolerance may be a reversible state, it is perhaps not surprising that 5–10 per cent of the population is autoimmune (Hugh McDevitt, Stanford University). McDevitt and his colleagues were responsible for establishing the striking association between a single amino acid in an MHC class I molecule and human autoimmune diabetes — a phenomenon that seems to have an equally striking animal counterpart in the non-obese diabetic (NOD) mouse. The introduction of the appropriate MHC transgene into a NOD mouse can confer protection against diabetes, as recently described by three independent groups, including that of Jacques Miller (Walter and Eliza Hall Institute, Melbourne) who discussed these experiments at the conference. The mechanism of protection is however still unknown (it is nothing so simple as the deletion of T cells bearing a potentially self-reactive receptor chain); and McDevitt's laboratory is now concentrating on other genes in the MHC of both mouse and man that may be associated with what he calls dominant suppression of autoimmunity. Not all genes in the MHC encode the class I and class II glycopro-

teins, as Townsend had already reminded us, and the quest for other genes associated with immune pathology may open new ways of exploring the immune response. A more immediately practical aim is to identify empirically those haplotypes associated with human autoimmune disease, with a view to monitoring individuals at risk for early signs of disease and possibly instituting therapy.

This raises the question of whether immunosuppression can be made specific to the autoreactive elements, leaving the rest of the immune system intact, for example by the use of peptides designed to mask specific T-cell determinants. Clearly the strategy can be successful only if autoimmunity depends on a strictly limited and identifiable subset of receptors. From that point of view it was discouraging to learn from Lawrence Steinman (Stanford University) that his most recent investigations into T-cell receptor usage in multiple sclerosis lesions in human brains reveal considerable variation, even from one lesion to another within a brain. Steinman did however report some success in reducing the severity of experimental allergic encephalomyelitis (EAE) in mice with either antibodies against the T-cell receptor or peptides binding strongly to the particular MHC molecule implicated in the allergic response.

Irun Cohen (Weizmann Institute) has been remarkably successful in vaccinating mice against peptide elements implicated in autoimmune arthritis and diabetes; and Leroy Hood (California Institute of Technology) was optimistic about the possibility of overcoming the problem of diversity of autoimmune responses by detailed analysis of the MHC and T-cell receptor haplotypes of susceptible individuals by means of molecular high technology. He finds that in autoimmune responses of mice to myelin basic protein, which leads to EAE, if both the MHC and the receptor haplotype are known the response repertoire is predictable. Human haplotypes however are more variable.

Exploiting diversity

Antibodies are at the evolutionary pinnacle of immune recognition mechanisms, adding to the germ-line and combinatorial diversity they share with the T-cell receptor the remarkable ability to undergo a high rate of somatic mutation. Mutation is confined to their variable region by a mechanism which Cesar Milstein (MRC Cambridge) has identified as *trans*-acting by transfecting B cells with 'passenger' immunoglobulin genes not subject to selection by antigen, and which seems to depend upon sequences downstream of the constant-region gene segment.

With marked hubris, Gregory Winter (MRC Cambridge) described a series of molecular manipulations for improving

upon the range and versatility of these outstandingly versatile molecules. His talk culminated with the suggestion that, by means of a new technique for directly selecting engineered bacteria for the production of proteins with specific binding properties, it may be possible to by-pass the antibody framework altogether and redesign (or at least reselect) antigen-binding molecules *de novo*.

The general idea that evolution may not be the ideal selection mechanism for therapeutic purposes was echoed by Jay Berzofsky (NIH), who described his attempts to design vaccines against diseases, such as AIDS, in which the immune response may make a large contribution to the pathology. This means that viral peptides for vaccination need to be designed to contain epitopes that will elicit neutralizing antibodies rather than those assisting (for example) the entry of viruses via Fc receptors into monocytic cells where they may flourish. Berzofsky quoted some evidence that neutralizing antibodies may help to prevent the infection by their mothers of babies with AIDS before birth. But neutralizing antibodies are present in infected adults, meaning that T-cell-mediated cytotoxicity may be a better route to eliminating the virus. He and his coworkers have recently reported their success in eliciting cytotoxic cells by injections of viral proteins in immunostimulatory complexes (ISCOMs) consisting of cholesterol and detergent in an animal model.

The conference ended with William Paul (NIH) on the basic science underlying the commonest immunopathology of all. Asthma and other allergies depend upon the triggering of mast cells by immunoglobulin E, and Paul's laboratory is trying to identify the mechanism whereby a B cell selects the heavy-chain gene segments encoding the epsilon isotype — an event in which interleukin-4 is strongly implicated. Because the isotype switch is one of many sophisticated immune mechanisms that are still not understood, a new technique he described for using the polymerase chain reaction to monitor its occurrence could well find widespread application.

It is perhaps the final irony of the molecular revolution in immunology that the techniques that have led to so great an increase in understanding in most aspects of the field have also provided its most impenetrable conundrum. Susumu Tonegawa (MIT), who helped to initiate the revolution with his unique contribution to understanding antibody diversity, summarized all that we now know about the $\gamma\delta$ T-cell receptor — its structure, ontogeny, patterns of diversity and expression, and even its apparent ability to recognize a newly defined non-classical MHC glycoprotein; in short, everything but its function.

Miranda Robertson

Order and chaos up in the air

Ian N. James

THE atmosphere undergoes a confusing mixture of random events, with tantalizing glimpses of order which are apt to dissolve even as you look at them. Ramamurthy *et al.* on page 314 of this issue¹ present case studies of one of the more ordered types of event observed in the atmosphere. Using a 50-MHz wind profiler and a 10-cm Doppler radar as well as traditional balloon-borne instruments, they have been able to measure the vertical structure of two exceptionally large-amplitude solitary waves which occurred over the United States. These propagated with little attenuation for over 1,000 km, leading to sudden changes of wind and intensity of precipitation as they passed by. The instrumentation available to the authors meant that they have been able to gain an unprecedented view of the vertical structure of these remarkable events.

Solitary waves are essentially nonlinear phenomena, and meteorology is perhaps the science *par excellence* of the nonlinear. Various conceptual models of the operation of nonlinearity have dominated thinking about fluid dynamics and hence atmospheric flow during this century. In the turbulent view, the nonlinear nature of atmospheric motions means that all scales of motion interact strongly with each other. Thus energy is scattered rapidly, irreversibly and essentially randomly throughout the entire range of possible modes of motion. Much of turbulence theory is based on such ideas of stochastic cascading of energy through the spectrum of motions.

The currently fashionable chaotic view suggests that a complex system such as the

atmosphere might be rather more ordered than the turbulent view would indicate. According to this view, the evolution of atmospheric flow, with its large number of degrees of freedom, is mainly confined to a sub-space of quite low, but fractal, dimension, embedded in the full phase space. The fractal structure of the sub-space means that despite its low dimensionality, the flow rapidly becomes unpredictable. Unavoidable uncertainties in the definition of the initial state become amplified exponentially as time passes, so that weather forecasts can be useful only for a limited time. On the longer time-scale, this view suggests that, for example, the detection of long-term changes in climate is extremely difficult, as the inherent unsteadiness and noisiness of any time series of atmospheric data will tend to mask any underlying trend.

A third insight into the operation of nonlinearity might be termed the 'coherence' view. The nonlinear nature of the governing equations can in some circumstances lead to considerable order and structure in the solutions, rather than to irregularity and disorder. The archetypal example is provided by the Korteweg-deVries equation (see box) which can be used to describe long-wavelength water waves and analogous atmospheric features. Solutions to the Korteweg-deVries equation can be constructed in which the nonlinear term exactly balances the other terms in the equation, resulting in a localized disturbance which propagates steadily without change of shape or amplitude — a solitary wave. The discovery of this phenomenon by J. Scott-

Russell^{2,3} in 1844 provides one of the more colourful episodes in nineteenth-century science. He describes a canal barge which suddenly stopped; its wake surged past the hull of the barge and propagated smartly along the canal as a localized solitary wave of elevation. Nothing daunted, our hero sprang onto his horse and galloped off in pursuit, and followed the wave for four miles before it (or the horse) finally expired.

In some circumstances, similar solitary waves can be observed in the atmosphere. The 'morning glory'⁴ enlivens an otherwise rather dismal tract of northeastern Australia with its spectacular roll clouds. These arise as an undular bore and its associated solitary waves propagates across the flat landscape. Although impressive in appearance, the morning glory has rather minor effects on the weather at ground level.

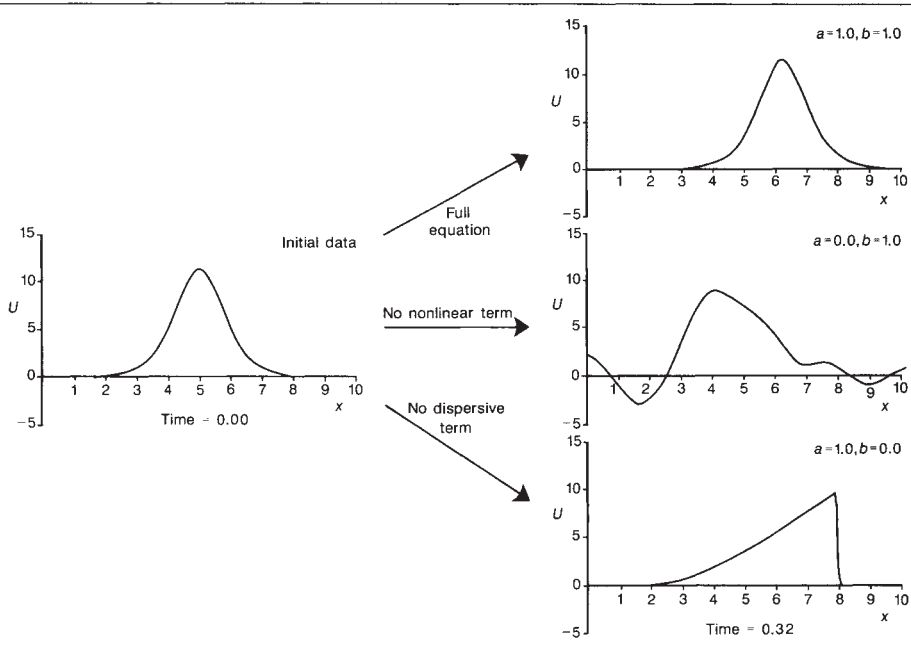
Atmospheric solitary waves require for their existence a vertical atmospheric structure that traps the disturbance in some kind of wave guide. Typically, a strong boundary-layer inversion can provide this. Similar, equally fleeting events, have now been documented in various parts of the world. But observing them is not easy. It requires the right instruments in the right place at the right time.

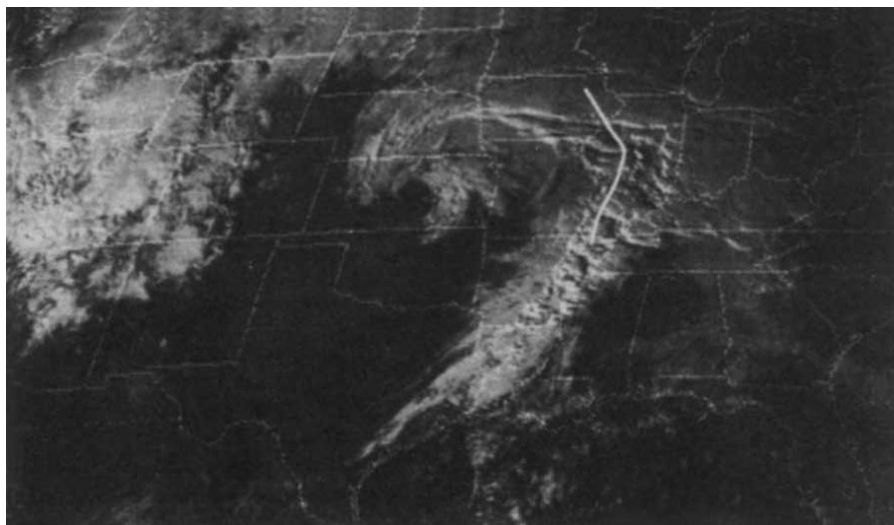
The solitary wave solutions of the Korteweg-deVries equation were derived as long ago as 1895 (ref. 5), but they were long regarded as a mathematical freak likely to be unstable and shortlived in the presence of any perturbations. Interest was renewed in the 1960s and early 1970s when numerical experimentation and then analytical work⁶ revealed that the solitary-wave solutions are in fact highly stable and robust. They have a pseudo-linear superposition property which means that colliding solitary waves

The Korteweg-deVries equation can be written in the form⁷:

$$\frac{\partial u}{\partial t} + a u \frac{\partial u}{\partial x} + b \frac{\partial^3 u}{\partial x^3} = 0$$

The second term is the nonlinear term and the third term is the dispersive term. If $a = 0$, we have a linear equation (right, middle); an initially localized disturbance (centre) will rapidly be smeared out as a result of the strong dispersive effect of the last term. If $b = 0$, the resulting nonlinear equation will steepen the gradients of any arbitrary initial disturbance to produce discontinuities or shocks (right, bottom). However, with both a and b non-zero, it is possible to choose an isolated initial disturbance such that these two competing effects exactly balance. Such a disturbance is a solitary wave; it propagates steadily without change of shape or amplitude (right, top). □





Satellite image of the frontal system associated with an intense low-pressure trough responsible for the solitary wave (white line) observed by Ramamurthy *et al.* on 5 January 1989.

interact strongly but emerge unscathed from the interaction. Similarly, a solitary wave can retain its identity for long periods even in the presence of a quite-large-amplitude, random, noisy background field. The events observed by Ramamurthy *et al.*¹ had just this character. The authors tracked their waves for over 1,000 km through all the chaotic mesoscale structure of an active depression system.

One of the interesting aspects of the observations is that the authors had instruments that could detect the structure of the solitary wave away from the surface. The data clearly show a solitary wave superposed upon an irregular background of fluctuations of substantial amplitude. The solitary waves the authors describe are truly spectacular, with huge vertical displacements of the streamlines. This large displacement means that they had a large effect on the weather, triggering rainstorms, as they passed by (see figure).

From the various case studies that have been published, it is beginning to seem that the occurrence of solitary wave events in the atmosphere is perhaps not so unusual. Just how frequent they are is a matter of conjecture, as many observations are fortuitous offshoots of field programmes directed at other goals. The conditions that allow the propagation of solitary waves can be defined reasonably well. But the circumstances that initiate them are largely unknown; they are unlikely to be further elucidated until better observations are available. If only it were possible to equip armies of enthusiastic ama-

teur meteorologists with cheap robust Doppler radar instruments. Amateur astronomers with home-made, but well-crafted telescopes provided just such a service to their science until very recently.

OPTICAL TRAPPING

Motor molecules in motion

James A. Spudis

ASSAYS *in vitro* that provide quantitative values for the velocity of movement of molecular motor molecules along either actin filaments or microtubules are now in common use. In this issue, reports by Block *et al.*¹ (page 348) and Ashkin *et al.*² (page 346) underscore the rapid increase in sophistication in measuring parameters of movement of such motors. In both studies (the first *in vitro*, the second *in vivo*) 'optical tweezers' were used to follow movements of small objects that have one or a few microtubule-based motor molecules attached.

Kinesin, a motor molecule that couples ATP hydrolysis to movement along microtubules³, is thought to drive vesicle transport and other forms of microtubule-based motility. Block *et al.*¹ devised an elegant way to watch a silica bead coated with one or only a few kinesin molecules move along a microtubule. They made use of a single-beam gradient-force optical particle trap to manipulate a bead and deposit it directly onto the microtubule, holding it there until the motor molecule 'finds' the microtubule track. This method increases the efficiency of the bead assay and allows measurements with approximately one kinesin molecule per bead.

The results with beads carrying large numbers of motor units, compared to those carrying a single or a few motor units, are striking on two counts — one is the distance travelled before the

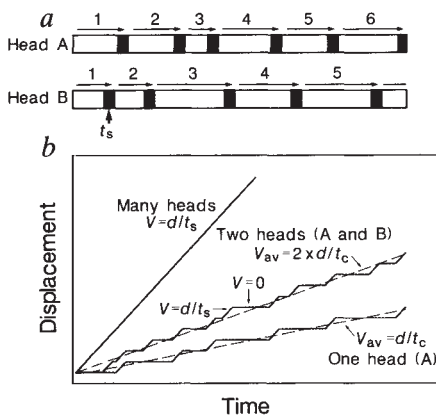
Two aspects of wider significance suggested by these new observations of solitary waves deserve highlighting. First, they give grounds for believing that certain mesoscale structures, which give rise to vigorous weather events, are highly predictable. 'Nowcasting' — the forecasting of events a few hours ahead on the smaller spatial scale — is seen as an important growth area in operational meteorology, with direct relevance to activities such as aviation. Recognizing predictable solitary-wave events could be a useful practical matter. Second, and more fundamentally, they suggest an alternative to the turbulent and chaotic views of nonlinear atmospheric flows which can be used in some circumstances. Elucidating the relationship of this alternative view to its rivals, and defining the situations when it is appropriate, is a challenging and fundamental problem in the modern study of meteorology. □

Ian N. James is in the Department of Meteorology, University of Reading, PO Box 239, Reading RG6 2AU, UK.

kinesin/bead complex escapes from the microtubule track, the other is the velocity of bead movement. When beads carry about 17 kinesin molecules, nearly all beads move 4 μm or more to the end of the microtubule to which they were attached. At about 1 kinesin molecule per bead, the beads move only about 1 μm , and then detach and diffuse away from the microtubule before reaching the end of the filament. Block and colleagues conclude that the most likely explanation is that the kinesin molecule releases very briefly (perhaps for about a millisecond) from the microtubule during each mechanochemical cycle. This 'stroke-release' model may be like that envisaged for the movement of myosin along actin, but the 'release' may be so brief that the bead usually does not have time to diffuse away from the microtubule. However, with some probability the bead will diffuse away from the microtubule during one of its release states, with the result that it leaves the microtubule before reaching the end.

This behaviour is extreme in the case of myosin, where low densities of motor units in a motility assay result in binding of actin filaments in the absence of ATP (the rigour state) but in their immediate release in the presence of ATP, with the result that the filament and myosin diffuse away from one another having experienced no detectable translocation at all⁴. Thus, myosin appears to be strongly

1. Ramamurthy, M.K., Collins, B.P., Rauber, R.M. & Kennedy, P.C. *Nature* **348**, 314–317 (1990).
2. Scott-Russell, J. *Brit. Assoc. Adv. Sci.*, 311–392 (1844).
3. Scott-Russell, J. *Proc. R. Soc. Edin.* 319–392 (1844).
4. Smith, R.K., Crook, N. & Roff, G. *Q. Jl R. met. Soc.* **108**, 937–956 (1982).
5. Korteweg, D.J. & deVries, G. *Phil. Mag.* **39**, 422–443 (1895).
6. Gardner, C.S., Greene, J.M., Kruskal, M.D. & Miura, R.M. *Phys. Rev. Lett.* **19**, 1095–1097 (1967).
7. Whitham, G.B. *Linear and Nonlinear Waves* (J. Wiley, New York, 1974).



a, Consecutive ATPase cycles of molecular motor heads. The average cycle time is t_c . Each cycle consists of a period when the motor is weakly bound (open rectangle) to its filament track, and a period when it is strongly bound (t_s , closed rectangle). b, Resultant displacement of a filament track with respect to its motors, on the same time scale as in a. Each motor undergoes a 'stroke' of step size d during the strongly bound state time. Each head operates independently and stochastically. The dashed lines show the average velocity (V_{av}). The average velocity of movement driven by a single head is limited to d/t_c , whereas that driven by many heads is limited to d/t_s . For a motor such as kinesin, where $t_s \approx t_c$, the velocity is independent of the number of motor units.

bound to an actin filament for a relatively short time (t_s) during the ATPase cycle time (t_c). Indeed, t_s/t_c for myosin has recently been measured at about 0.05 (ref. 5). For kinesin^{1,6} this ratio may be closer to 1 (perhaps about 0.5 per head, assuming the two heads of the kinesin molecule move in a hand-over-hand coordinated fashion along a microtubule).

What is the velocity of the kinesin/bead complex carrying one or so kinesin molecules compared to that carrying many? Strikingly, Block *et al.*¹ show that the velocities are very similar. This confirms earlier work by Howard *et al.*⁶, who used kinesin molecules attached to a glass surface and watched microtubules translocate across them. These results with kinesin contrast markedly with observations on myosin, where the velocity decreases significantly as the number of myosin motor units working on an actin filament is reduced⁵.

The difference in behaviour may be accounted for by the apparent differences in the t_s/t_c ratios exhibited by these two very different motors. According to one model, which is undoubtedly an oversimplification of the transduction process, a stroke of a given 'step size' d occurs for each ATPase cycle that the motor unit undergoes. This model predicts that the average velocity of movement along a single motor unit should be limited by d/t_c , whereas the velocity of movement of a filament interacting with a large number of motor units would be limited by d/t_s (see figure). For kinesin, preliminary data sug-

gest that t_s is close to t_c , so that $d/t_s \approx d/t_c$, whereas for myosin $d/t_s \approx 20 \times (d/t_c)$. Although this explanation is not the only way to view the existing data, it is exciting that the *in vitro* motility assays are being developed to a further level of sophistication that allows testable predictions to be made.

The difference in behaviour between kinesin and myosin may reflect their different functions *in vivo*. Conventional myosin works in the form of a thick filament, which contains many motor units. In this situation, no single motor unit needs to be tightly bound to the actin filament for a long time. In contrast, kinesin presumably attaches specifically to, and moves, cytoplasmic organelles that are too small to accommodate many motors. So here the t_s/t_c ratio must be quite large, or the organelle would detach and float away without any significant translocation having occurred.

How do the forces in these systems *in vitro* compare to estimates of those operating *in vivo*? Ashkin *et al.*² describe the use of the optical trap to estimate force generation of organelle transport *in vivo*. They combined this novel noninvasive approach with a useful model system, the giant amoeba *Reticulomyxa*, in which, it has been suggested⁷, mitochondrial movements are driven by cytoplasmic dynein, another microtubule-based motor. Ashkin *et al.* trapped an actively translocating mitochondrion at high power, and then quickly reduced the power to see whether it could escape. The force of any escaping organelle has a value between that corresponding to the upper and lower limits

PROTON TUNNELLING

Sorting chickens from eggs

J. A. Krumhansl

"ROUNDAABOUT the accredited and orderly fact of every science," according to the philosopher William James, "there ever floats a sort of dust cloud of exceptional observations, of occurrences minute and irregular and seldom met with, which it always proves more easy to ignore than to attend to." Elsewhere in this issue, M. I. McMahon *et al.* (Nature 348, 317–319; 1990) present the final, unavoidable evidence that demolishes the conventional view of the phase transition in hydrogen-bonded ferroelectric materials, compounds that spontaneously acquire an electric dipole below a critical temperature.

The report concerns the comparative effects of bond lengths, the bistable hydrogen-bond geometry and the substitution of deuterium for hydrogen in KDP, the prototypical ferroelectric, on the material's ordering temperature. The holy canon in the field is that the change in

fixed by the stopping and escape power settings. The authors estimate that, on average, about 1–4 motor molecules are associated with each *Reticulomyxa* mitochondrion. And the value that they estimate for the force generated by the *Reticulomyxa* motor is similar to values estimated *in vitro* for other motors^{8,9} — in general, molecular motors appear to be capable of generating force in the range of a piconewton.

It seems that in the not-too-distant future a relatively simple assay will be available for widespread use to measure force produced by molecular motors. Expression and molecular genetic manipulation of both kinesin and myosin have been achieved. So we are not far from the time when mutational analyses can be coupled to resultant alterations in force-velocity curves, as determined *in vitro* with highly purified proteins. □

James A. Spudich is in the Departments of Cell Biology and Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

1. Block, S. M., Goldstein, L. S. B. & Schnapp, B. J. *Nature* **348**, 348–352 (1990).
2. Ashkin, A., Schütze, K., Dziedzic, J. M., Euteneuer, U. & Schliwa, M. *Nature* **348**, 346–348 (1990).
3. Vale, R. D., Reese, T. S. & Sheetz, M. P. *Cell* **42**, 38–50 (1985).
4. Toyoshima, Y. Y., Kron, S. J. & Spudich, J. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7130–7134 (1990).
5. Uyeda, T. Q. P., Kron, S. J. & Spudich, J. A. *J. molec. Biol.* **214**, 699–710 (1990).
6. Howard, J., Hudspeth, A. J. & Vale, R. D. *Nature* **342**, 154–158 (1989).
7. Euteneuer, U., Koonce, M. P., Pfister, K. K. & Schliwa, M. *Nature* **332**, 176–178 (1988).
8. Oiwa, K. & Takahashi, K. *Cell Struct. Funct.* **13**, 193–205 (1988).
9. Kishino, A. & Yanagida, T. *Nature* **334**, 74–76 (1988).

ordering temperature brought about by deuterium substitution is attributable to the change in isotopic mass and the way this alters the probability of tunnelling between the two stable hydrogen-bond geometries. The new neutron-diffraction data, placed alongside results from the literature, show convincingly that this cannot be the case.

Although certainly of importance to understanding hydrogen-bonded ferroelectrics, there are broader implications to this work: proton-tunnelling units are embedded in a variety of molecular compounds — ice and biomolecules such as rhodopsin are examples in which cooperative proton tunnelling plays an important role. In this broader context it is apparent that one must deal not only with the proton but also with its associated matrix, so that a wide range of behaviours is then possible.

A priori, one would expect that, in

KDP, both the surrounding lattice configuration and the effective potential for proton motion (or localization), all coupled together, would be involved in determining the process of proton ordering, and (in addition) the development of polarization upon transition from the disordered to the ordered, ferroelectric state. Conceptually this amounts to considering two subsystems, the proton tunnelling subunit, and the host lattice molecule, which are coupled. As to the behaviour of this system, the problem in general is akin to the chicken and the egg: the proton's (deuteron's) environment and dynamics are both determined by and act on the host, and vice versa. Obviously, a wide range of outcomes is possible, and the theorist is tempted to oversimplify in modelling. In the history of this topic, reasonably influenced perhaps in the 1960s by the observation that deuterium substitution for hydrogen causes a large increase in the transition temperature T_c , it was assumed that proton tunnelling completely dominates the transitional physics — that is, that the host is unchanged chemically or structurally and is simply a rigid, static environment. Obviously, in the general case that might be drastically wrong physically although instructive for modelling.

But as McMahon *et al.* point out, there has been a mist of uncooperative facts: Ubbelohde noted, in 1939, that H/D substitution changes the geometry of the hydrogen bonds; in 1968, Buyers *et al.* found no 'pseudospin' (or two-state) softening of the cooperative proton tunnelling in KDP, which had been predicted by theory; and a number of authors have noted from studies of the effect of pressure that the change in T_c correlates well with the length of the bonds linking the PO_4 tetrahedra. Other anomalies have been noted.

The growing availability of high-quality materials, convenient high-pressure cells and precision neutron-diffraction data has now allowed the characterization of the entire system, host and proton, with a sufficient number of experimental control parameters to probe and separate the host from the tunnelling physics. One thing is clear from the detailed conclusions McMahon *et al.* draw: the chicken-and-egg, host-and-tunnelling system is not separable. New theories are needed for these ferroelectrics. Indeed, the situation in condensed-matter physics and materials science generally has become more demanding of the theorist: dominant, universal mechanisms are the exception as we look realistically at coupled, complex systems. That is why a new look at the wider class of proton-tunnelling systems is in order. □

J. A. Krumhansl is in the Department of Physics, Cornell University Clark Hall, Ithaca, New York 14853, USA.

Search for the missing links

R. M. Denton

THOSE involved in unravelling the complex series of events which follow the binding of insulin to cell surface receptors, and which lead to its many intracellular effects, will find cause for excitement in the report by Dent *et al.* on page 302 of this issue¹. The work, which comes from Cohen's group in Dundee, may take us a step towards understanding at the molecular level how insulin regulates cell metabolism.

It is now generally accepted that binding of insulin to its receptor results in the immediate stimulation of the tyrosine kinase activity of the intracellular domain

Serine protein kinases that are activated in cells incubated in the presence of insulin

Acetyl-CoA carboxylase kinase
ATP-citrate lyase kinase
Casein kinase II
Insulin receptor serine kinase
Microtubule associated protein-2 kinase (MAP-2 kinase)
Myelin basic protein kinase
Protein kinase C
Raf-1
Ribosomal protein S6 kinase I (70K)
Ribosomal protein S6 kinase II (90K)

of the receptor, and that this is essential for all subsequent events^{2,3}. Among the events is the activation of an array of protein kinases which phosphorylate specific serines (and threonines) within potentially regulatory proteins^{3,4}. These protein kinases are presumably responsible for bringing about the observed increases in serine/threonine phosphorylation of a number of cellular proteins, including the ribosomal protein S6, acetyl-CoA carboxylase, ATP-citrate lyase and the insulin receptor itself. But one of the paradoxes about insulin is that it also brings about the dephosphorylation of other proteins; indeed, many of the best known effects of the hormone, for example on glycogen and triacylglycerol metabolism, occur in this way. So it is of great interest that Dent *et al.* have come up with evidence of an insulin-activated serine protein kinase which can phosphorylate and activate the major protein phosphatase involved in the dephosphorylation of glycogen synthase and phosphorylase kinase. Activated serine/threonine protein kinases may therefore be essential links between the occupied insulin receptor and all the intracellular changes in protein phosphorylation which follow.

Previous studies by Cohen's group showed that, in rabbit muscle, dephosphorylation of the enzymes regulating glycogen metabolism (glycogen synthase, phosphorylase and phosphorylase kinase)

is catalysed by a form of one of the main cytoplasmic protein phosphatases (PP1), in which the catalytic subunit (of relative molecular mass 37,000; M_r 37K) is complexed with a 160K subunit (known as the G subunit). Studies *in vitro* have demonstrated that the G subunit is involved in both directing the phosphatase to the glycogen particle and regulating the activity of the catalytic subunit. It now seems that phosphorylation on one serine (site 2) by cyclic AMP-dependent protein kinase causes dissociation of the catalytic subunit and diminishes its activity, whereas phosphorylation on another serine (site 1) results in an increase in its activity against glycogen synthase and phosphorylase kinase. Moreover, Dent *et al.* show that injection of fasting rabbits with insulin results in both a twofold increase in the phosphorylation of site 1 as well as a threefold increase in the activity of a kinase capable of phosphorylating this site.

The table lists some of the serine protein kinases reported to be activated within a few minutes of incubating cells in the presence of insulin. These kinases act on a wide range of protein substrates, but direct evidence that phosphorylation actually causes changes in activity of these targets has not usually been forthcoming⁵. An exception is the acetyl-CoA carboxylase kinase found in fat cells, which leads to a stimulated form of acetyl-CoA carboxylase and so to increases in fatty acid synthesis⁶. Acetyl-CoA carboxylase also shares with the G subunit the characteristic that phosphorylation on different serines can give rise to opposing changes in activity.

The kinase studied by Dent *et al.* may be related to the 90K S6 kinase II (see table), as the two have similar substrate specificities and are inactivated by treatment with protein phosphatases^{1,7}. This second property suggests the kinases are part of a kinase cascade. In the case of S6 kinase II from *Xenopus* oocytes, it has been found that partial reactivation can be achieved by another insulin-activated serine protein kinase, known as MAP-2 kinase, which is found in many different cell types. Activation of this kinase in insulin-treated cells is associated with increased phosphorylation of the kinase on both threonine and tyrosine residues⁷. However, MAP-2 kinase does not seem to be a direct substrate of the insulin receptor tyrosine kinase, and changes in activity seem to correlate more closely with changes in threonine phosphorylation⁷. In any case, the main insulin-activated S6 kinase in most cells is the smaller '70K' enzyme, which although also apparently part of a kinase cascade is not apparently a

substrate for MAP-2 kinase (ref. 8; L. M. Ballou and G. Thomas, personal communication).

The hunt is now on for the missing 'kinase-kinases' and for the kinases which may precede them in a cascade — but such putative kinases are only likely to be present in cells in very small amounts. None of those shown in the table is known to be a direct substrate of the insulin receptor itself, although Raf-1 and the insulin receptor serine kinase may be associated with it^{9,10}. The principal missing link in the insulin-signalling system remains the event (or events) which follows the activation of the insulin receptor tyrosine kinase and which may initiate these cascades of serine protein kinases.

The study of Dent *et al.* points to other questions that need answers. For example, are the time courses of the activation of the rabbit muscle kinase and the phosphorylation of site 1 on the G subunit sufficiently fast to explain the rapid effects of insulin on glycogen synthesis? Which phosphatase dephosphorylates the G

subunit and how is that phosphatase regulated? And finally — and most importantly — how does insulin bring about the dephosphorylation of other proteins in cells, such as triacylglycerol lipase and pyruvate dehydrogenase (a mitochondrial enzyme), which are acted upon by different phosphatases? □

R. M. Denton is in the Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, UK.

1. Dent, P. *et al.* *Nature* **348**, 302–308 (1990).
2. Zick, Y. *Crit. Rev. Biochem. molec. Biol.* **24**, 217–269 (1989).
3. Denton, R.M. *Adv. Cyclic Nuc. Prot. Phosph. Res.* **20**, 293–401 (1986).
4. Czech, M.P., Klarlund, J.K., Yagaloff, K.A., Bradford, A.P. & Lewis, R.E. *J. biol. Chem.* **263**, 11017–11020 (1988).
5. Haystead, T.A.J. *et al.* *J. biol. Chem.* **265**, 16571–16580 (1990).
6. Borthwick, A.C., Edgell, N.J. & Denton, R.M. *Biochem. J.* **270**, 795–801 (1990).
7. Maller, J.L. *Biochemistry* **29**, 3157–3166 (1990).
8. Price, D.J., Gunsalus, J.R. & Avruch, J. *FASEB J.* **4**, Abst. 2206 (1990).
9. Blackshear, P.J., Haupt, D.M., App, H. & Rapp, U.R. *J. biol. Chem.* **265**, 11231–11234 (1990).
10. Smith, D.M. & Sale, G.J. *Biochem. J.* **256**, 903–909 (1988).

RELATIVITY

Opening a can of wormholes

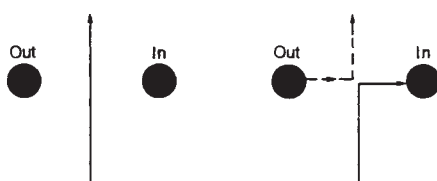
C.J.S. Clarke

PERHAPS riding on the success of the films 'Back to the Future *i*' (*i* = 1,2,...), time machines are creeping back into the news with a paper¹ from Kip Thorne and colleagues suggesting that the problems they pose concerning our notion of history may not be as severe as had been feared.

The basic problem raised by time travel has been discussed since the appearance of Kurt Gödel's model universe in 1949. General relativity does not necessarily admit the idea of a general time-coordinate applicable to the whole of the Universe; instead, time is a local matter, applicable only in a limited region of space and for limited durations. There is thus no reason why there should not be particles, or even objects, in the Universe whose history in space-time has the form of a circle. At any point on the circular history there is a locally defined time with a well-defined direction of future and past, but the history returns to its starting point. Can this really happen?

It was possible to dismiss Gödel's example because it described a rotating universe that does not correspond to the one we observe. Harder to dismiss was Kerr's solution for a rotating gravitational source, which was shown by Carter in 1968 to contain closed timelike curves when the rotation was fast. This could well represent the final state of a collapsing star whose mass was too great for it to form a neutron star, and so in principle such objects might exist in the Universe. The physical details of the process are still

uncertain, however. Computer simulations by Nakamura and Oohara suggest that in the course of collapse, a star would shed matter and so reduce its spin below the critical value for closed timelike curves. But if the star failed to get rid of



Non-unique evolutions in the presence of a time machine (from ref. 1). A billiard ball aimed initially between the two mouths of a wormhole time machine (black circles) could pass through without being affected at all (left), or it could be deflected into the entrance of the time machine by a version of itself, at a later stage of its history, emerging from the machine's exit (right). There is evidence that there exist initial conditions for which there is no consistent evolution.

enough spin, it would constitute a genuine time machine. Not only would there be circular histories in the depths of the stellar remnant, but also it would be possible to send projectiles into the remnant and have them emerge before they went in ('before' being assessed by reference to a time coordinate in a region sufficiently far from the star for general relativity to affect it only weakly).

The problems created by such a possibility are similar, whatever the mechanism for producing the time machine. The

current work¹ is in fact mainly concerned with a different and, at first sight, less plausible mechanism (previously discussed in News and Views² by Friedman) based on wormholes³ — thin tracts of space connecting two regions that could be widely separated, as seen from the outside. It turns out that these can rather easily produce closed timelike curves that pass through the wormhole. Since 1983 it has been known⁴ that a wormhole will collapse as soon as it is created unless it is maintained in a state in which the energy of the vacuum is shifted by quantum processes to be negative. So there remains grave doubt, first as to whether a wormhole could ever be formed in the first place, and second, whether there is any feasible means for creating enough negative energy in this way to hold it up.

The new paper tackles the basic problem with time machines, which, as every science-fiction reader knows, is that you could go back in time and kill your mother before you were born. Formulated in this way, the problem becomes hopelessly embroiled with the nature of free will; but it is easy to imagine purely mechanical devices⁵ that give rise to the same sort of paradox. For these one can ask a well-defined physical question. If we set up a particular physical situation before the time machine is constructed, in which there is a device programmed to execute apparently paradoxical behaviour, what, according to the laws of physics, will happen?

We can think of this as a problem in the theory of differential equations. The world is described by some collection of physical fields, obeying partial differential equations, with initial values prescribed throughout all space before the time machine comes into existence. If the initial situation is really paradoxical, then these equations will have no solution. Thorne *et al.* show that for the simplest case, when the fields in question consist just of a single non-interacting scalar field, the field equations do have a solution for the case of the wormhole time machine, and in each case this solution is unique. They also suggest, by consideration of classical mechanical problems in the presence of time machines, that for more realistic fields there may be some initial situations for which there are no solutions and some for which the solution is not unique (see figure). They propose a Principle of Self Consistency to rule out the occurrence of such conditions.

The proof that there exist unique solutions to at least simple field equations in wormhole space-times is particularly interesting for cosmology, because it goes against the long established wisdom of the subject, which runs as follows: to do physics you need to predict (to solve equations uniquely from initial conditions); mathematicians tell us that, in general, equa-

RÉSUMÉ

Lunar reflections

IN the quest for ever-higher resolution in astronomy, there is a fundamental practical problem: a restricted aperture size on a telescope limits the resolution achievable. That is why radio-astronomers resort to very-long-baseline interferometry, with multiple-component telescopes such as the Very Large Array in New Mexico and MERLIN in the UK. But even these devices are restricted to the space available on our planet. So, for good measure, why not use the Moon as the radio mirror, ask T. Hagfors and colleagues (*Astrophys. J.* **362**, 308–317; 1990)? The Moon's surface is neither smooth nor very reflective. But incorporated into an interferometer array, it might just help resolve details of bright radio bursts on Jupiter or maser emission from Orion.

Aged introns

A CLASS of introns that interrupt certain genes of eukaryotes may have been present in the common ancestor of two major groups of bacteria, according to D. Shub and colleagues (*Cell* **63**, 417–424; 1990). Until now, the only known examples of introns in genes of eubacteria were three found in different genes of bacteriophage T4, a virus that infects gram-negative bacteria. These three were classed as 'group I' introns, because of the presence of characteristic conserved sequences. The question has been whether these introns are present in bacteria and eukaryotes as a result of their ancient origin or horizontal gene transfer. Shub *et al.* have discovered a group I intron in a gene of bacteriophage SPO1 which can infect only gram-positive bacteria, adding force to the idea that the introns have an ancient origin and have existed for at least 1,500 million years (the estimated time since gram-negative and gram-positive bacteria diverged from their common ancestor). A further twist is that in all cases the eubacterial introns interrupt genes involved in DNA synthesis: the authors speculate that introns have been conserved because they are part of a system for regulating synthesis of viral DNA.

Pelvic parameters

DETERMINING gender from skeletal parameters is one of the preoccupations of anatomists and palaeontologists. A refinement from N. Milne is a set of handy discriminant functions (*J. Anat.* **172**, 221–226; 1990) based on principal components analysis of eight dimensions of the human pelvis, derived from 62 unsexed skeletons in a medical collection. The analysis separated the pelvis into two clear groups based on gender, as confirmed by the answers to a separate, subjective questionnaire from a panel of anatomists.

tions like those in physics can only be solved in space-times where every history leading into the past from a given event hits the initial surface; time machines (and naked singularities) violate this condition; so time machines are incompatible with physics, and we ought to forget about them. It appears that this argument is wrong. Although 'in general' time machines may prevent prediction, there can still be particular cases such as the wormhole space-times where this is not the case.

While this may get the wormholes off the hook, there remains a worry about the other more realistic cases, different space-times and nonlinear fields. Computer simulations of collapsing stars are progressing fast. Suppose we found an initial configuration that does collapse to produce a time machine, of the sort where some initial field conditions do not have solutions. Should we be content at this stage to wheel in a Principle to solve our

problems? Most theoreticians would want to go further and ask, what goes wrong with an attempted solution? The most likely outcome, which is suggested by this work, would seem to be that these cases are unstable, and that the back-reaction of the physical fields on the space-time would destroy the time machine. If this proved to be the generic case, then we would be able to forget about time machines, not because they were embarrassing, but because there were good physical mechanisms for preventing them. □

C. J. S. Clarke is in the Department of Mathematics, University of Southampton, Southampton SO9 5NH, UK.

1. Friedman, J.L. *et al.* *Phys. Rev. D* **42**, 1915–1930 (1990).
2. Friedman, J.L. *Nature* **336**, 305–306 (1988).
3. Morris, M.S., Thorne, K. & Yurtsever, U. *Phys. Rev. Lett.* **61**, 1446–1449 (1988).
4. Lee, C.W. *Gen. Relativity Gravitation* **15**, 21–30 (1983).
5. Wheeler, J.A. & Feynman, R.P. *Rev. mod. Phys.* **21**, 425–434 (1949).

EVOLUTION

Living fast and dying young

J. S. Jones

BIOLOGY used to be all about lists: lists of cranial nerves, lists of species and — soon to come — the biggest list of all, the DNA sequence of the human genome. Animal behaviour in particular was until recently just a vast catalogue of anecdotes on sex, life and death. A recent meeting* might have been (but was not) called 'Explaining David Attenborough': just why do animals do all the cute or disgusting things we see on television? It turns out that some simple rules govern the lives of creatures as different as flowers and apes — and might even help to make sense of some of the eccentric behaviour of DNA.

Life-history theory can be summed up as 'nobody gets a free lunch'; no creature can simultaneously be good at growth, survival and reproduction. There is always a trade-off, and success in one walk of life is paid for by failure in another. What lifestyle to adopt is determined by where an organism lives and by what its competitors are doing. Compromise is everywhere. In *Drosophila*, stocks selected for long adult life have low juvenile survival as larvae and grow for longer than normal (L. Partridge, University of Edinburgh). In plants, too, there is a trade-off of the resources allocated to flowers and to fruit (D. Charlesworth, University of Chicago). The conflict between components of fitness extends over generations: parents must bargain their investment in progeny against the loss of future reproductive success due to the expense of raising their first-born. The divergence in the

interests of parents and children reduces optimal family size, and explains why birds are willing to let their youngest chicks starve if conditions are less than ideal (C. Godfray, Imperial College London).

Sex itself involves trade-offs. It is expensive, and males may die during the risky business of attracting and retaining a mate. Sometimes the expense makes it worth cheating; to mate and flee in the hope that, by cuckolding a more responsible father, reproductive success will emerge on the cheap. In blue-gill sunfish up to 90 per cent of adult breeding males behave in this way, but the costs of cuckoldry are high as such males are often exposed to predators. The two strategies — a safe and faithful life versus a seductive but dangerous one — are alternative solutions to the problem of trading sex against survival; and the proportion of each is stable within a population (M.L. Gross, University of Toronto).

There is a deeper conflict in sex, that between somatic tissue (a source of metabolites) and the gametes (a metabolic sink). This may help to explain aspects of the shift from unicellular to multicellular animals. In *Volvox* and its relatives, somatic cells do not appear until the plants reach a size of about 100 cells. As colonies get larger, relatively more somatic cells are needed to maintain the germ cells. Mutants which reduce the number of somatic cells in a colony have a disproportionate effect on the number of germ cells which can be supported; sex is a costly habit (G. Bell, McGill University). In other creatures, too, reproduction eats

*The Evolution of Reproductive Strategies. Royal Society, London, 17–18 October 1990.

into an energy budget which might otherwise be used to repair somatic cells. Promiscuous male and fecund female *Drosophila* die young. In hermaphrodite plants there is a trade-off between male and female function, which means, for example, that self-fertilizing plants allocate less energy to producing pollen than do outcrossers. Ageing itself may be an unwelcome side-effect of sex. As reproduction is so costly, it is no longer possible to maintain somatic cells for indefinite periods. The idea of a 'disposable soma' explains why embryonic stem cells in culture are more or less immortal, whereas somatic cells soon senesce (T.B.L. Kirkwood, MRC, Mill Hill, London).

It is easy to see trade-offs in nature but harder to establish their genetic basis. Environmental effects often confound the issue. Individuals who find a particularly favourable habitat may succeed both in living longer and having more offspring: as Partridge pointed out, cars and houses are both expensive, but the biggest cars are to be found outside the biggest houses. There are few tests of the genetics of life history. Artificial selection on one fitness component may lead to an opposing change in another, although the effect is sometimes subtle. In *Drosophila*, for example, populations selected for rapid increase in numbers show no change in adult survival or fecundity, but they do evolve changes in larval growth (L.D. Mueller, University of California, Irvine). Perhaps the best evidence of the genetic basis of a trade-off comes from the faithless fish — cuckolding males carry genes which lead to earlier maturation and smaller size than those guarding a nest.

Another way to study the evolution of life histories is by comparing related species. Here there are the usual problems of disentangling similarity through shared descent from that arising from adaptation. P. Harvey (University of Oxford) showed how new methods of comparing evolutionary trees reveal a close fit between relative age at maturity and life expectancy in creatures as different as chipmunks and elephants. Even their parasites share these trends. New statistical techniques for looking at ratios (E.L. Charnov, University of Utah) uncover some intriguing — and largely unexplained — differences in lifestyle. Birds, for example, live five times longer after maturity than do fish. Many differences in life history relate to differences in metabolic rate, although quite why is not yet understood.

The consistency of some patterns of life history suggests that they reflect fundamental rules of life, rules which may also help to explain patterns of DNA sequence evolution. Much DNA in eukaryote genomes seems to evolve more in its own interests than in that of its carriers. There are many trade-offs in the life history of such sequences, including the conflict

between reproduction and repair common to the functional DNA. Rapidly reproducing transposable elements in *Drosophila* are more often deleted from their insertion sites than are the quiescent and stable repeated sequences which make up much of the eukaryote genome. There are even instances of molecular cuckoldry: 'defective' transposons in *Drosophila* — which, like the cheating fish, are much smaller than their relatives — reproduce by hijacking the replication mechanism of complete elements infecting the same cell. Just as in fish, incomplete sequences suffer a high mortality and can outnumber those investing more metabolic effort in reproduction.

The conflicting interests of repeated sequence DNA and its carriers is anala-

gous to that between parent and offspring, as the maintenance and replication of the excess DNA reduces the metabolic energy available for reproduction of functional sequences. Animals such as salamanders, which live in stable environments and have low fecundity and low metabolic rate, carry many repeated sequences. In contrast, the DNA of bacteria and of mitochondria — which are metabolically active and must replicate rapidly in a highly competitive environment — are free of such extraneous material. It could be that ecological theory will end up telling us as much about molecular biology as it does about animal behaviour. □

J.S. Jones is in the Department of Genetics and Biometry, University College London, London NW1 2HE, UK.

ORION NEBULA

Star formation's tangled web

Glenn J. White

STAR formation is a complex process in which giant interstellar clouds containing millions of solar masses of gaseous hydrogen collapse under gravity to form individual stars. The physical processes involved are complex and remain poorly understood. Recent measurements at radio wavelengths by Farad Yusef-Zadeh (*Astrophys. J.* **361**, L19–L22; 1990) towards the Orion Nebula, have revealed an intricate network of ionized filaments which have arisen from the star formation process. These may be due to the interaction of material ejected from the mass-losing stars at the centre of the Orion nebula, or be related to shock fronts around ionized (H II) regions. In addition

to these small-scale features, a number of other more extended ionization fronts show a cone-like structure, which seem to be related to the overall structure of the region. In particular, what can they tell us about whether shocks can destroy or energize star-forming clouds, and how do they promote or inhibit star formation?

The Orion nebula is the closest site to us at which high-mass stars (spectral classes O or B having masses greater than 5 solar masses ($M_{\odot}$)) are forming. Its relatively close proximity of 470 parsecs (1 parsec = 3×10^{16} m) and location 150 parsecs below the galactic plane, make it an ideal place to study the mechanisms of star formation. The Orion star-formation complex

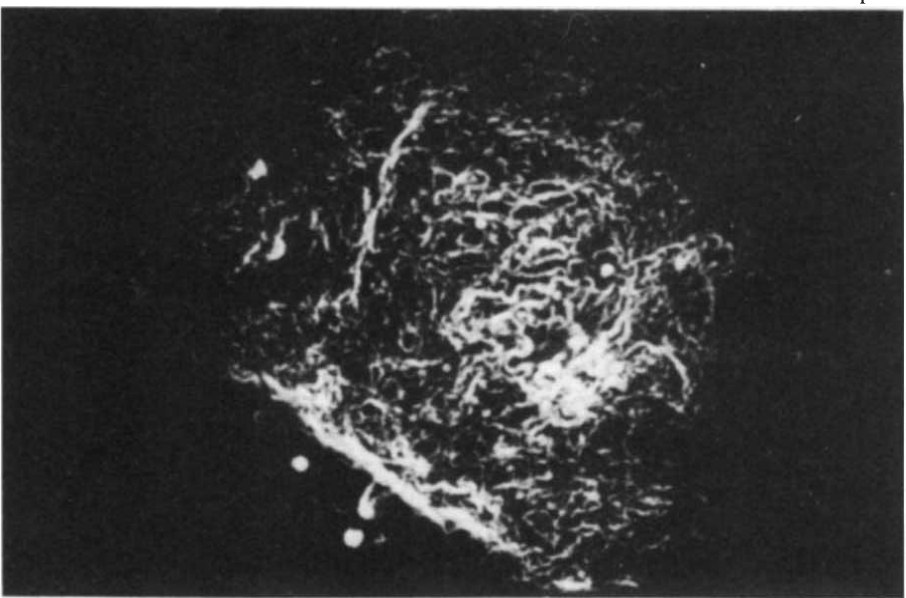


FIG. 1 Radiophotograph of the central regions of the Orion Nebula which has been enhanced with a sharp mask to show the widespread filamentary structures present. In the lower left hand corner, the linear feature shows the main ionization bar, which is clearly resolved into many individual narrow features, which merge together. To the right of the centre, the streamers are seen to be clustered together, surrounding the positions of the Trapezium Cluster.

F. Yusef-Zadeh

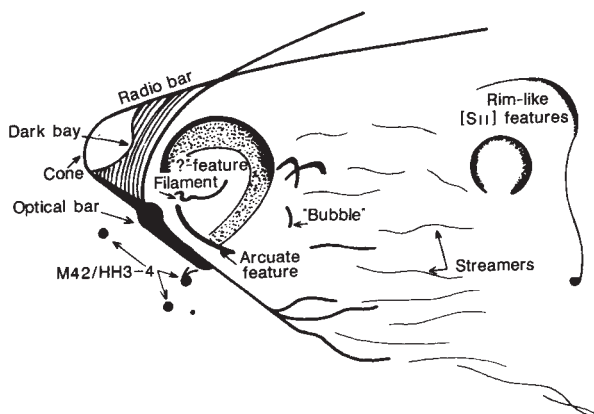


FIG. 2 A schematic view of the Orion Nebula viewed face-on, illustrating some of the prominent structures and features which are present (From F. Yusef-Zadeh *Astrophys. J.* **361**, L19–L22; 1990).

consists of a large envelope (over $10^5 M_{\odot}$) of neutral hydrogen gas ($H I$), several large molecular-hydrogen clouds (H_2) of about $10^5 M_{\odot}$, $H II$ regions and dark clouds, stretching over almost 20 degrees on the sky. Observations at individual molecular-line wavelengths (J. Bally *et al. Astrophys. J.* **312**, L45–L48; 1987) show that the material is concentrated into large-scale filaments or sheets, containing dense molecular cores in which star formation may be occurring.

At visible wavelengths the outstanding feature is the naked eye-nebula Messier 42, illuminated by a young group of OB stars just several million years old, the Trapezium Cluster. These are the brightest members of the I Orion OB stellar association, which contains around 60 O6–B2-class stars, which were probably formed through a sequential star-formation mechanism (B.G. Elmegreen & C.J. Lada *Astrophys. J.* **214**, 725–741; 1977). In this process, the first-generation stars are formed when the edge of a molecular cloud is 'triggered' into gravitational collapse by a violent external event, such as a shock front from a nearby supernova explosion. These first-generation stars may then interact, via stellar winds or radiation pressure, with their placental material, disturbing it and inducing a new generation of star formation.

Observations at optical wavelengths of star-forming regions are often hampered by the absorption and scattering of light by interstellar dust. Most information comes from the molecular-line observations of the neutral gas at millimetre wavelengths, or from radio-wavelength continuum studies of the ionized gas. Yusef-Zadeh made his new radio continuum observations with the Very Large Array (VLA) aperture-synthesis telescope in New Mexico at 20 cm wavelength. The high-angular-resolution (1.8×1.6 arcsecond beamsizes) images so made cover a relatively large field of view (about one quarter of a square degree), allowing the large-scale morphology of small-scale structure in the gas to be studied.

This is related to the more global mechanisms of star formation.

The radiophotograph (Fig. 1) shows a spectacular network of features centred on the Trapezium cluster. The web-like filamentary structures indicate that stellar winds and ionizing radiation from the central Trapezium Cluster have significantly disturbed the nearby material, creating narrow ionization fronts and strings. The effects of an expanding shell that has resulted from the collision of a stellar wind with mater-

ial having a strong density gradient will lead to the growth of hydrodynamic Rayleigh–Taylor instabilities in the gas. These cause widespread fragmentation, cooling and the formation of discrete condensations, similar to those reported in the new observations.

On the larger scale, the ionized gas delineates the edge of a cone (Fig. 2), with its apex lying close to an opaque region known as the optical bay. Ionized gas, driven by the internal pressure of the $H II$ region, appears to be impeded from further expansion by this dense region. The data are consistent with the following picture of star formation in the Orion Nebula. The main area of present-day high-mass star formation contains a cluster of about ten $15\text{--}25 M_{\odot}$ protostars or pre-main sequence objects lying close to the front edge of a dense molecular cloud core just behind the centre of the visible nebula. Immediately surrounding these stars, gaseous material is subjected to considerable heating and photo-ionization, which excavates a cavity around the star as it clears away this placental material. Outside this, the effects of radiation from the Trapezium cluster give rise to many of the optically visible features of the Orion Nebula, and in particular show how recently formed stars may interact with surrounding material leading to further induced star formation.

The Orion Nebula has a very complex and intricate structure, with many dense filaments, clumps and sheets interspersed with low-density ionized gas, bubbles and cavities. The interaction between the newly formed stars and the surrounding material clearly has a strong influence on the nebula's appearance and evolution. Ultimately star formation will use up or disperse most of the material in the cloud, leaving behind an extensive stellar association, similar to those seen in many other parts of the Galaxy. $\square$

Glenn J. White is in the Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK

That inward eye

DOCTORS and physiologists frequently need to look inside the human body. All their current methods have drawbacks: X-rays can be damaging, NMR tomography is expensive, ultrasound is cumbersome and not always informative. So why not use light? After all, human tissue is translucent, as can easily be seen by shining a torch through your hand. The snag, of course, is that human tissue scatters light strongly. Simply standing a patient in a light beam like a slide in a projector, would give a hopelessly foggy image.

So Daedalus is refining the idea. He proposes to illuminate the patient, not with continuous light, but with a rapid sequence of sub-picosecond pulses from a mode-locked laser. Optically useful, image-forming light will traverse the patient directly, and will emerge first. Scattered light will take a circuitous route, suffering one or more deflections, and will be delayed. By 'time-gating' the emerging light pulses so as to select only their leading edges, a clear visible image should result, free of obscuring scatter. The brevity of the gating time will stretch the art of high speed optical switching to its limits, but the thing looks feasible.

So DREADCO's technicians are scrutinizing naked laser-illuminated volunteers through lithium niobate optical-switch spectacles, while tweaking the gating electronics for maximum subjective transparency. As with X-rays, bones and foreign bodies appear as opaque shadows. Arteries, veins and organs show up according to their effective colour, or fluoresce in the laser light.

By aiming the laser into the patient from the front, and gating the back-reflected light, he could be examined in depth, radar fashion. A fixed time delay between the emission of each light pulse and the brief opening of the optical-shutter spectacles would define a precise out-and-return distance, isolating a narrow slice of the patient's interior for scrutiny. By moving the source back or forward, the slice could be scanned through the patient in a form of optical tomography.

At last the doctor will be freed from the tedious interpretation of screens and photographs. Instead, he will examine and scan through his patient directly. Wearing optical-shutter spectacles and aiming a pulsed laser torch, he will be able to peer at the beating heart, study the movement of a joint or the flexing of a muscle, press on suspect areas to see how the organs beneath respond, check that pills have been correctly swallowed or that an implant is safely in place, select the best approach for surgery, and so on. A patient wearing white cotton or nylon clothes that scatter but hardly absorb light, may not even have to undress.

David Jones

Handling excess nitrates

SIR—Because modern agriculture practices are now widely recognized as a significant source of nitrate pollution, it is time to consider ways in which land can be managed to absorb the lost nitrate before surface water becomes contaminated.

Traditionally, floodplains in Britain were not vigorously farmed, but land drainage now allows these zones to be ploughed up or managed more intensively as grassland. The unforeseen consequence of this activity is that nitrate-contaminated ground water, which was once allowed to drain slowly through the floodplain, is now conducted rapidly across the floodplain into the stream or river via ditches or tile drains. Nitrate-rich subsurface water that enters an 'undrained' floodplain soil zone is likely to lose most of its nitrate load through processes such as denitrification and assimilation^{1,2}.

Our work in the upper Thames basin, southern England, shows that a grass-covered floodplain retains the ability to reduce significantly the nitrate concentration of ground water throughout the winter (mean loss of nitrate of 82% for the winter of 1989–90). Our work also demonstrates that the reduction processes seem to operate for different types of surface vegetation, in that the loss of nitrate has occurred in both poplar and grass-covered floodplains. For a major runoff event in January 1990, the nitrate concentration of ground water increased by about 400%, but the grassed floodplain still maintained a nitrate-buffering capacity close to its mean level. The reduction of nitrate in the floodplain has not been accompanied by the excessive production of ammonium ions, which could be one of the indirect consequences when nitrate-rich ground water enters a surface water system.

We conclude that floodplains need to be preserved in (or returned to) their undrained state as these areas sustain a potential to reduce nitrate concentrations in ground water throughout the year. By optimizing the role of floodplains as buffer zones, and thereby preventing ground water rich in nitrate from entering the stream, the quality of drinking water will be improved, and the risk of excessive macrophyte growth (and the consequent problems associated with the decay of those plants in the autumn) will be reduced.

We believe that an undrained 'green corridor' on either side of streams and rivers (especially in headwater catchments) should be preserved as the first line

of defence for aquatic ecosystems. Cleaner surface water in headwater catchments will provide greater flexibility on river use, and will have significant benefit for the management of estuarine and marine environments.

NICHOLAS HAYCOCK
TIM BURT

School of Geography,
University of Oxford,
Mansfield Road,
Oxford OX1 3TB, UK

Making sense of information

SIR—Greene's study¹ of colonial ospreys (*Pandion haliaetus*) in Nova Scotia is one of six that strongly supports the 'information centre' hypothesis². Greene suggested that members of a colony of these birds obtained information on the location of ephemeral schools of fish from other members. Greene's conclusions have since been challenged³ and here I interpret some of his data in a different way.

My colleagues and I have also examined osprey social foraging in Nova Scotia. We found that colonial ospreys at the foraging area directly observed where conspecifics were foraging and used that information to find food themselves, a strategy called local enhancement⁴. To determine why our conclusions were different from Greene's, I visited the same area as that studied by Greene, as well as analysing the data in his thesis⁵.

During the period when there were pollock (*Pollachius virens*, schooling species) in Cow Bay and Cole Harbour estuaries, Greene made most of his observations at Cow Bay⁵, which was visible from the osprey colony. Thus, the observation that ospreys fly in the direction from which pollock were delivered within 10 min of delivery can be most simply explained by a local enhancement strategy. For deliveries of pollock, the data collected at Cow Bay should therefore be excluded from the analysis, and only then can an assessment be made of whether information was transferred.

Local enhancement can also explain the response to ospreys that made conspicuous flight displays on discovering a school of fish. As all ospreys at the colony flew to the displaying bird, they must have been able to observe it foraging, otherwise how did these birds 'know' to initiate hunting near where the fish had been caught? Although the display directed attention to the school, information was individually acquired.

Greene reported¹ that the other two schooling species in the area, smelt (*Osmerus mordax*) and alewife (*Clupeus*

harengus), were highly localized in a few freshwater spawning sites and, therefore, distributed predictably. If these sites were being used by ospreys (as seems likely), then with the resulting preferred foraging destinations, angles of departure were bound to retrace those of arriving birds. To analyse these data appropriately would require omitting departure and arrival angles corresponding to these sites.

Foraging at freshwater spawning sites also provides an alternative explanation of how ospreys located distant schools solely on the basis of incoming flight directions. Finding the same school requires that it remain stationary long enough for the successful forager to return to the colony, plus up to 10 min before departure (using Greene's methodology), plus the time required to return to the school. This seems unlikely for distant schools because of the long time period required. Perhaps what Greene thought were distant schools were in fact stable spawning sites. Indeed, the spawning site 5 km from the colony⁵ corresponds to the maximum recruited foraging distances he reported¹.

I conclude that Greene's results can be interpreted in other ways. Reanalysis is required before concluding that the osprey colony he observed functioned as an information centre.

STEPHEN P. FLEMMING

Department of Biology,
Queen's University,
Kingston, Ontario K7L 3N6, Canada

GREENE REPLIES—I have reanalysed my data, excluding all fish caught within sight of the Cow Bay colony or in the ocean between a point of land about 1.5 km to the east of the colony and one 1.75 km south. The percentage of fish caught out of sight of the colony varied among fish species: 34.0% of pollock, 50.8% of alewife, 83.7% of smelt and 66.0% of flounder. Ospreys that departed within 10 min of another bird returning from out of sight with a pollock tended to fly in the direction from which the successful bird had arrived. By contrast, birds leaving the colony within 10 min of another bird returning from out of sight with a winter flounder did not tend to fly in that direction.

Alewife and smelt are anadromous fish, spawning during April to June in two freshwater systems that flow into Cow Bay. If ospreys from the colony were foraging at stable spawning aggregations, then a correlation between arrival and departure directions could arise even in the absence of the transfer of information about distant food sources. Alewife and smelt can be found anywhere along the streams and in lakes up to 5 km from the colony. However, ospreys returning to the colony with these fish tended to return from only two compass directions (roughly 280° and 320°, corresponding to the directions of the mouths of the streams). On

1. Haycock, N.E. & Burt, T.P. in *Hydrological Research Basins and the Environment* (eds Hooghart, J.C., Posthumus, C.W.S. & Warner, P.M.M.) 335–339 (TNO Committee on Hydrological Research, The Hague, 1990).

2. Pinay, G. & Décamps, H. *Regulated Rivers* 2, 507–516 (1989).

any given day during the spawning season, successful foragers returned from both stream systems. Non-foraging ospreys at the colony showed a strong tendency to fly up the stream system from which a forager had just returned. This indicates that departing foragers were sensitive to the stream system in which the previous smelt or alewife had been caught.

Thus it appears that ospreys at this colony made use of information from ospreys fishing nearby (local enhancement) and, to a lesser extent, information from birds returning from foraging locations out of sight of the colony (information centre). But information transfer about distant food supplies is not a general feature of aggregations of breeding ospreys (refs 3,6; S. Flemming, personal communication). As I pointed out in my paper¹, as ospreys do not follow successful birds back to foraging areas, information transfer at colonies will be useful only in situations where the foraging areas are not

too far from the colony. At other colonies, ospreys have to fly further simply to reach the foraging areas (ref. 3; S. Flemming, personal communication) than ospreys at the Cow Bay colony. Taken together, these results suggest that information transfer at osprey colonies is probably a fortuitous consequence of the local geography of fishing locations and the spatial and temporal distributions of available fish species, rather than a general factor promoting aggregated breeding.

ERICK GREENE

*Division of Biological Sciences,
University of Montana,
Missoula,
Montana 59812, USA*

1. Greene, E. *Nature* **329**, 239–241 (1987).
2. Ward, P. & Zahavi, A. *Ibis* **115**, 517–534 (1973).
3. Hagan, J. & Walters, J. *Auk* **107**, 506–521 (1990).
4. Thorpe, W. *Learning and Instinct in Animals* (Methuen, London, 1956).
5. Greene, E. thesis (Dalhousie Univ., Halifax, 1982).
6. Poule, A. *Ospreys: a Natural and Unnatural History* (Cambridge University Press, 1989).

Sea-level flips

SIR—In the Mediterranean Sea, evaporation exceeds precipitation and river discharge. This drives a relatively fresh surface inflow through the Strait of Gibraltar (Fig. 1) and a slightly weaker outflow of saltier water near the bottom^{1–3}. Geodetic levelling along the Spanish coast⁴ shows an associated drop in mean sea level from Atlantic to Mediterranean of about 0.15 m.

Recent monthly mean sea-level data (Fig. 2), however, show that this sea-level drop is not steady but changes seasonally by about 0.1 m. We believe that understanding this variation is relevant to the broader question of whether the Strait of Gibraltar is a good climatic 'choke point' at which long-term changes in air-sea interaction over the whole Mediterranean might be detectable.

Mathematical models^{5,6} show that there are two possible exchange states. One occurs if the interface between inflowing and outflowing waters at the western end of the Mediterranean is shallower than about 90 m. The surface inflow then leaves the strait as a fast, shallow, jet with

a sea-level drop of about 0.20 m along the north shore due to a combination of the Bernoulli effect and friction. The exchange is said to be 'maximal' as it is the maximum consistent with the density difference and hydraulic constraints in the strait. 'Submaximal' exchange occurs if the interface just inside the Mediterranean is deeper than about 90 m; the inflow is then slower and deeper, with a mean sea-level drop of only about 0.09 m. For maximum exchange the interface depth in the strait is fixed, whereas for submaximal exchange it can rise and fall in response to changes in the volume of dense water.

It is remarkable that the difference of 0.11 m between the sea-level drop predicted for the two states is close to the changes seen in Fig. 2, suggesting that the state flips seasonally. Moreover, if each state occurs for about half the year (Fig. 2), the expected average sea-level drop of 0.15 m agrees with the result from levelling⁴. The implied flips could be related to the formation of dense Mediterranean water in the winter. This would raise the level of the interface and push the system into maximal exchange, where it might stay for the first half of the year until the reservoir of dense water has been drained down below about 90 m again^{7,8}.

In our recent survey⁸, we found other evidence for state flips. In particular, during the 1985–86 Gibraltar experiment⁹, it appeared from direct measurements in the strait^{10,11} that the inflow was slow and deep in October 1985, but fast and shallow in April and early May 1986 with some supporting evidence⁸, from the inferred interface height at the eastern end of the strait, for a flip towards the end of December 1985 (Julio Candela, personal communication). Figure 2 is partly consistent with this, though it suggests that the exchange was already maximal by December 1985.

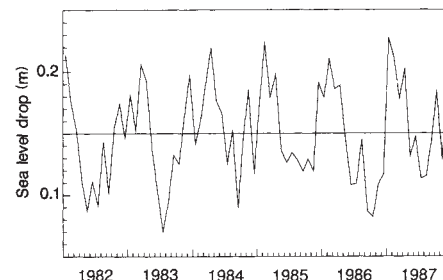


FIG. 2 Monthly mean sea level drop from Cadiz to Malaga from January 1982 to December 1987, using data obtained from the Permanent Service for Mean Sea Level. Data are plotted with a mean of 0.15 m (ref. 4). Adding suitably scaled atmospheric pressure to give the total pressure difference causes an r.m.s. change of less than 0.01 m.

The sea-level differences shown in Fig. 2 are monthly means. Meteorologically driven fluctuations, with periods of several days, could lead to flips within a month, accounting for the monthly average values in Fig. 2 that are intermediate between the two end-states. On the other hand, our analysis of sea-level fluctuations within the strait from September 1981 to September 1982 led us to the unsettling conclusion that the exchange was maximal during that period¹². This is inconsistent with Fig. 2, which shows a period of submaximal exchange in the middle of 1982.

Present models thus fail to account for all the features in the observations and we cannot be completely sure that the hydraulic state changes seasonally. If, however, the exchange is indeed submaximal for part of the year, interannual and long-term variability in air-sea fluxes affecting the formation of dense water should be seen at the Strait of Gibraltar as a changing interface level⁸. More geodetically referenced sea-level and other oceanographic data from the Strait of Gibraltar are needed.

CHRIS GARRETT
KEITH THOMPSON
WADE BLANCHARD

*Department of Oceanography,
Dalhousie University,
Halifax, Nova Scotia B3H 4J1, Canada*

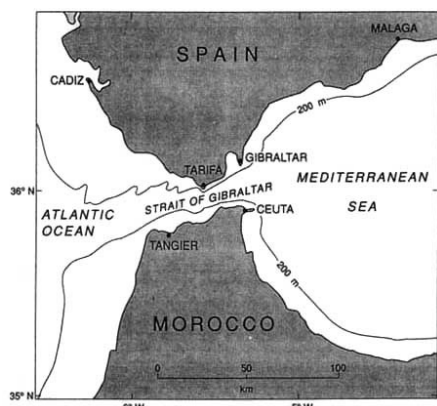


FIG. 1 The Strait of Gibraltar and its environs.

1. Deacon, M. *Scientists and the Sea 1650–1900* (Academic, New York, 1971).
2. Lacombe, H. & Richez, C. in *Hydrodynamics of Semi-Enclosed Seas* (ed. Nihoul, J.C.J.) 13–73 (Elsevier, Amsterdam, 1982).
3. Bryden, H.L. & Stommel, H.M. *Oceanol. Acta* **7**, 289–296 (1984).
4. Levallois, J.J., & Maillard, J. in *Report on the Symposium on Coastal Geodesy 300–330* (Munich, 1970).
5. Farmer, D.M. & Armi, L. *J. Fluid Mech.* **164**, 53 (1986).
6. Bormans, M. & Garrett, C. *J. phys. Oceanogr.* **19**, 1543–1557 (1989).
7. Bormans, M., Garrett, C. & Thompson, K.R. *Oceanol. Acta* **9**, 403–414 (1986).
8. Garrett, C., Bormans, M. & Thompson, K.R. in *The Physical Oceanography of Sea Straits*, ed. Pratt, L.J. (NATO/ASI Series, Kluwer, Dordrecht, 1990).
9. Bryden, H.L. & Kinder, T.H. *Oceanol. Acta SP* 29–40 (1988).
10. Wesson, J.C. & Gregg, M.C. in *Small-Scale Turbulence and Mixing in the Ocean* (ed. Nihoul, J.C.J.) 201–212 (Elsevier, Amsterdam, 1988).
11. Armi, L. & Farmer, D.M. *Prog. Oceanogr.* **21**, 1 (1988).
12. Garrett, C., Akerley, J. & Thompson, K.R. *J. phys. Oceanogr.* **19**, 1682–1696 (1989).

Meteoritic silicon carbide: pristine material from carbon stars

Roy S. Lewis*, Sachiko Amari** & Edward Anders*

* Enrico Fermi Institute and Department of Chemistry, University of Chicago, Chicago, Illinois 60637-1433, USA

All five noble gases in interstellar silicon carbide grains have grossly non-solar isotopic and elemental abundances that vary with grain size but are strikingly similar to calculated values for the helium-burning shell of low-mass carbon stars. Apparently these grains formed in carbon-star envelopes, and were impregnated with noble gas ions from a stellar wind. Meteoritic SiC provides a detailed record of nuclear and chemical processes in carbon stars.

THE first hint of red-giant material in meteorites was a tiny xenon component¹ (7×10^{-5} the total Xe) that showed the characteristic signature² of the s-process (neutron capture on a slow timescale): strong enrichment of the even-numbered isotopes 128, 130, 132 over the other six isotopes. Because the s-process is known to occur in red giants (specifically in AGB, or asymptotic giant branch, stars), the xenon must have originated in a red giant. Further clues began to emerge when the carrier of this Xe-S component was isolated³. It turned out to be SiC, containing highly anomalous Si, C and N^{4,5}, as well as Kr-S⁶⁻⁸ and Ne-E(H), one of two Ne components strongly enriched in ²²Ne (ref. 9), which is released at high temperatures and is attributed¹⁰ to the decay of ²²Na (half-life 2.6 yr).

There are apparently several populations of SiC, because the Si and C isotopic compositions of individual grains fall into several discrete clusters^{5,11}, the Xe-S/Ne-E ratio varies with grain size^{3,5}, and the proportions of ⁸⁰Kr and ⁸⁶Kr vary with release temperature of the gas on heating^{7,8}. The variations of ⁸⁰Kr and ⁸⁶Kr reflect branching of the s-process at their radioactive progenitors, ⁷⁹Se and ⁸⁵Kr (ref. 7). These branchings depend sensitively on neutron density and temperature in the s-process region^{7,12}, and thus can provide clues about the stars in which the Kr was formed. What we would like to learn from a detailed study of the SiC is: (1) the number of parent stars; (2) their properties—mass, composition and evolutionary stage; (3) their connection, if any, to the early Solar System; (4) the nuclear, physical and chemical processes in the stars and in interstellar space.

The anomalous noble gases in SiC are diluted by other, more normal noble gas components, presumably located in other phases such as spinel or diamond. We therefore prepared two highly purified SiC samples and separated them into grain size fractions. Most of the normal gases disappeared, but a small amount remained and is presumably located in SiC. The isotope correlations of these two components strikingly resemble predictions for two regions of an AGB star^{13,14}: the He-burning shell, containing s-process products, and the envelope, containing less-processed material. Here we report the principal experimental results and first-order inferences; the astrophysical implications are discussed in greater detail in the companion paper by Gallino *et al.*¹⁴. A full account of the experimental data will be published elsewhere.

We isolated SiC and other interstellar grains from two ~80-g samples (L and K) of the Murchison C2 chondrite, using variants of procedures developed in our laboratory^{3,5,15}. The parent SiC samples LQA (1.8 p.p.m.) and KJ (5.8 p.p.m.) were then separated by centrifugation into five and nine grain size fractions, respectively, LQB to LQF and KJA to KJI. All except KJI were analysed for noble gases by stepped heating¹⁶. Both sets were >90% pure SiC according to scanning electron microscopy/energy-dispersive X-ray analysis, but the L-series samples were smaller and had lower noble gas concentrations, probably implying losses of gas-rich material. Accordingly, we focus mainly on the K series.

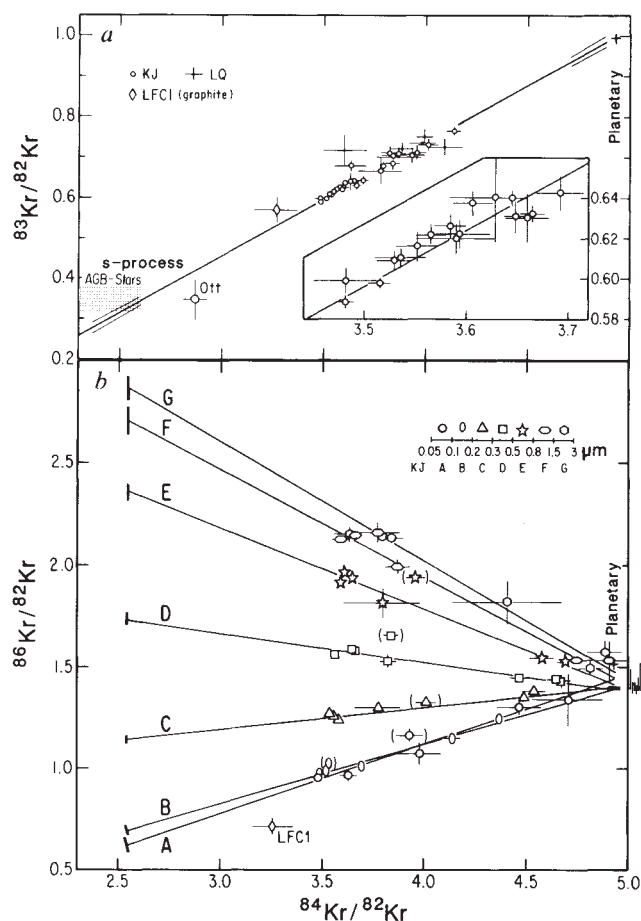


FIG. 1. Analysis of Kr by stepped heating. *a*, Three-isotope plot of ⁸²Kr, ⁸³Kr and ⁸⁴Kr, which are all on the main s-process path. The linear array shows that they are binary mixtures of a normal component close to planetary Kr and an s-process component in the range of theoretical compositions estimated by Gallino *et al.*¹⁴. *b*, Three-isotope plot of ⁸²Kr, ⁸⁴Kr and ⁸⁶Kr. ⁸⁶Kr, located at a branch point of the s-process path, forms a separate linear array for each grain size fraction, but with progressively larger y intercepts. ⁸⁶Kr/⁸²Kr ratios of s-process components (short, heavy line segments at ⁸⁴Kr/⁸²Kr = 2.55) increase with grain size from A to G, implying that Kr in the coarser grains formed at higher neutron densities. The error bars on the far right are for samples KJA to KJG at ⁸⁴Kr/⁸²Kr = 4.928.

† Present address: McDonnell Center for Space Sciences, Washington University, St Louis, Missouri 63130-4899, USA.

TABLE 1 Noble gases in SiC size fractions from Murchison chondrite

Sample	Size (μm)	$^3\text{He}/^4\text{He}$ $\times 10^{-4}$	$^{20}\text{Ne}/^{22}\text{Ne}$	$^{38}\text{Ar}/^{36}\text{Ar}$	$^{84}\text{Kr}/^{82}\text{Kr}$	$^{130}\text{Xe}/^{132}\text{Xe}$	$\left(\frac{^{80}\text{Kr}}{^{82}\text{Kr}}\right)_s$	$\left(\frac{^{86}\text{Kr}}{^{82}\text{Kr}}\right)_s$	^4He	Gas concentration ($10^{-8} \text{ ml g}^{-1}$)			
			^{20}Ne	^{36}Ar	^{82}Kr	^{132}Xe				^{22}Ne	^{36}Ar	^{82}Kr	^{130}Xe
KJA	0.05–0.1	1.32 (18)	0.9075 (10)	0.1927 (5)	3.702 (24)	0.3532 (10)	0.0597 (67)	0.621 (44)	4.82×10^6	5,130	547	3.82	8.28
KJB	0.1–0.2	2.00 (10)	0.8933 (3)	0.1919 (1)	3.603 (6)	0.3602 (4)	0.0586 (16)	0.695 (24)	7.09×10^6	9,300	685	5.97	13.74
KJC	0.2–0.3	1.48 (7)	0.6345 (3)	0.1907 (2)	3.693 (10)	0.3604 (6)	0.0526 (27)	1.145 (19)	8.28×10^6	13,400	678	4.79	9.12
KJD	0.3–0.5	1.16 (6)	0.4582 (5)	0.1935 (1)	3.769 (8)	0.3558 (5)	0.0490 (25)	1.733 (23)	9.36×10^6	19,600	675	4.26	5.96
KJE	0.5–0.8	0.81 (4)	0.2977 (5)	0.2051 (2)	3.756 (11)	0.3465 (10)	0.0451 (21)	2.363 (34)	10.58×10^6	28,400	575	4.23	3.20
KJF	0.8–1.5	0.56 (3)	0.2001 (2)	0.2210 (5)	3.776 (21)	0.3291 (22)	0.0365 (37)	2.710 (71)	10.28×10^6	35,800	497	4.40	1.72
KJG	1.5–3	0.53 (5)	0.1556 (2)	0.2313 (6)	3.915 (21)	0.2716 (28)	0.0336 (72)	2.872 (62)	6.93×10^6	28,800	371	3.86	0.56
KJH	3–5		0.0979 (19)	0.2090 (23)	4.500 (87)	0.1606 (52)	0.055 (58)			6,290	218	1.76	0.21
Solar ²¹		1.42	13.7	0.188	4.988	0.1643							
He-Shell, Range*			0.05–0.084	0.57–1.25	2.2–2.6	0.39–0.44							
He-Shell, Typical		0	0.0808	0.74	2.55	0.485†							

Numbers in parentheses are the 1σ uncertainties in the last one or two digits.

* The first line gives the range for AGB-star He shells²²; the second line gives the specific values adopted for resolution of the data.

† Xe-S, from SiC samples measured in this study.

Krypton-S: multiple sources

To resolve the Kr data, we present them on three-isotope plots with a common denominator (Fig. 1), where mixtures of two components lie on a straight line connecting these components. First, we must find the endmembers for ^{82}Kr , ^{83}Kr and ^{84}Kr , which are not affected by branching in the s-process. They indeed form a linear array (Fig. 1a), suggesting that they are binary mixtures of Kr-S in the lower left and some isotopically 'normal' component (solar, meteoritic or atmospheric) in the upper right. Although we purposely omitted all gas fractions obtained below 1,000 °C, which are dominated by normal Kr from extraneous phases, error-weighted linear regression lines¹⁷ for various combinations of the remaining data (with or without the LFC1 graphite sample¹⁸) consistently pass through normal Kr, specifically 'planetary' or 'AVCC' (average carbonaceous chondrite) Kr (ref. 19). Just as consistently, they pass through the range of theoretical Kr-S values¹⁴, although somewhat off the mid-point. As a compromise, we have adopted $(^{84}\text{Kr}/^{82}\text{Kr})_s \equiv 2.55$, which for our regression line gives $(^{83}\text{Kr}/^{82}\text{Kr})_s = 0.329 \pm 0.011$. We shall use this value for resolution of the remaining isotopes.

The Kr-S point of Ott *et al.*⁷ at $^{84}\text{Kr}/^{82}\text{Kr} = 2.87 \pm 0.06$ does not lie on our regression line. After showing that a correlation similar to that in Fig. 1a missed the theoretical range, they obtained the above Kr-S value by extrapolating a $^{84}\text{Kr}/^{82}\text{Kr}$ – $^{136}\text{Xe}/^{82}\text{Kr}$ plot of four temperature fractions, omitting the largest fraction with 44% of the Kr-S. But the latter correlation seems to have been fortuitous, as our data do not show it although they are much more accurate and comprise 36 fractions from 15 samples compared to the 5 fractions from 1 sample of lower purity of Ott *et al.* (its Kr-S concentration was 300 times lower than that in our parent sample, KJ).

The $^{86}\text{Kr}/^{82}\text{Kr}$ ratios (Fig. 1b) increase steadily with grain size. The size fractions form separate linear arrays fanning out from a common origin and extrapolating to progressively larger $^{86}\text{Kr}/^{82}\text{Kr}$ ratios for the pure s-process component at $(^{84}\text{Kr}/^{82}\text{Kr})_s = 2.55$. In most cases, the 1,200 °C fractions (in parentheses) lie above the regression line, perhaps implying the presence of another mineral with a distinctive Kr component and lower but sharper release temperature than SiC. Because these aberrant fractions contain only 4–14% of the Kr, it seems permissible to omit them.

The lines for the four samples with the smallest errors (B–E) converge on a point slightly below planetary Kr, corresponding to $^{86}\text{Kr}/^{82}\text{Kr} = 1.401 \pm 0.011$ at $^{84}\text{Kr}/^{82}\text{Kr} = 4.928$ (black bar in Fig. 1b). This is the 'normal' Kr component in SiC, and apparently differs slightly from planetary, solar or atmospheric Kr, at least for ^{86}Kr . The remaining three lines are consistent with this value within their larger errors. The complete range of $(^{86}\text{Kr}/^{82}\text{Kr})_s$ ratios is 0.62–2.87 (Table 1), compared to 0.78–1.57 found by Ott *et al.*⁷, and implies a range of neutron densities^{12,14}.

A similar trend with grain size was first observed in the Murchison LQ series⁸, but was less clear owing to the larger errors. In retrospect, it seems that the trend of Ott *et al.* also was due to grain size: $(^{86}\text{Kr}/^{82}\text{Kr})_s$ increases with combustion temperature (which presumably increases with grain size), except for a reversal between 900 and 1,080 °C. The extrapolated ratios of Ott *et al.* are, however, incorrect (although possibly covered by their larger errors), being based on $(^{84}\text{Kr}/^{82}\text{Kr})_s$ and $(^{86}\text{Kr}/^{82}\text{Kr})_{\text{normal}}$ ratios that are too high.

The spread for $^{80}\text{Kr}/^{82}\text{Kr}$ is much smaller, but an analogous extrapolation gives $(^{80}\text{Kr}/^{82}\text{Kr})_s$ ratios (Table 1) that decrease with increasing $(^{86}\text{Kr}/^{82}\text{Kr})_s$, as expected in the s-process. These values are lower than those of Ott *et al.* but within the error

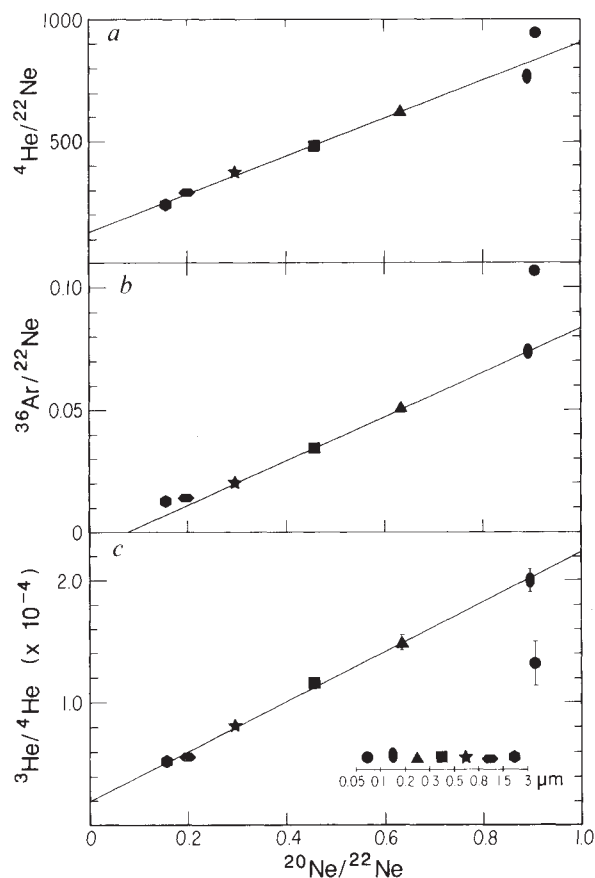


FIG. 2 Three-isotope plots of a, $^4\text{He}/^{22}\text{Ne}$ and $^{20}\text{Ne}/^{22}\text{Ne}$ and b, $^{36}\text{Ar}/^{22}\text{Ne}$ and $^{20}\text{Ne}/^{22}\text{Ne}$. The linear correlation corresponds to mixtures of an anomalous component (G) in the lower left and a normal component (N) offscale to the upper right. c, Four-isotope plot of $^3\text{He}/^4\text{He}$ against $^{20}\text{Ne}/^{22}\text{Ne}$.

limits of the latter, and likewise suggest a low temperature in the s-process region ($\sim 1.5 \times 10^8$ K), consistent with a ^{13}C but not a ^{22}Ne neutron source^{13,14,20}.

At first sight it seems surprising that Kr-S isotope ratios correlate with SiC grain size, because isotopes reflect s-process conditions in the hot interior of the star whereas grain size reflects condensation conditions in the cool atmosphere. But the latter depend on mass loss rate, stellar luminosity and ionization by radioactive nuclei, all of which, like s-process conditions in the interior, are ultimately functions of stellar mass and metallicity. A more detailed analysis¹⁴ suggests that the SiC comes from one or more AGB stars of mass $1-3 M_{\odot}$ and metallicity Z of 0.01–0.02.

Comparison with AGB stars

Having measured all five noble gases in SiC, we can try to compare elemental as well as isotopic abundances with those in AGB stars. The principal isotope ratios of all noble gases lie between He-shell and solar²¹ values (Table 1), suggesting that all gases are diluted by a 'normal' component, which must be corrected for. In several cases, three-isotope plots are linear, showing that only two components are present (Figs. 1, 2*ab* and 4). Designating these two components N (for normal) and G (for AGB He shell), we can resolve them by the standard formula: $\alpha = (R - R_N)/(R_G - R_N)$, where R is the observed isotope ratio I/J, R_G and R_N are the corresponding ratios of the pure components, and α is the fraction of isotope J belonging to the G component.

For R_G , we have used the typical He-shell or s-process values in the last line of Table 1. For R_N , we should, in principle, use AGB-star envelope values, in view of evidence presented below that the N component comes mainly from this source. But the envelope compositions vary over a considerable range¹⁴ and, as all these compositions are ultimately derived from solar composition by addition of stellar nucleosynthesis products, we have simply used solar composition, effectively reassigning these additions to the G component. The choice of solar or meteoritic composition as an endmember is supported by stepped-heating data, which lie on correlation lines passing close to solar-meteoritic values (Figs 1 and 5). Abundances of the two components are generally accurate to better than a factor of 1.5, because the measured isotopic compositions lie well away from the plausible ranges of the endmembers (Table 1). The only exception is for Ar in KJA-KJD, where the isotope ratios are very close to solar.

Helium required a different approach because $^3\text{He}/^4\text{He}$ gave nonlinear trends on three-isotope plots, implying the presence of a third (apparently cosmogenic) component. We therefore resolved ^4He on the basis of the elemental ratios from Fig. 2*a*—($^4\text{He}/^{22}\text{Ne}$)_G = 193 ± 8 and ($^4\text{He}/^{20}\text{Ne}$)_N = 800 ± 26 —and the $^{22}\text{Ne}_G$ and $^{20}\text{Ne}_N$ values found from the $^{20}\text{Ne}/^{22}\text{Ne}$ ratio.

We now compare the SiC data with calculated values^{13,14} for the He shells of AGB stars of low mass ($1-3 M_{\odot}$) and metallicities ranging from 1×10^{-3} to 2×10^{-2} (Fig. 3). All data are normalized to solar values²¹; the elemental abundances are also normalized to ^4He —a nuclide that is enhanced by less than a factor of three during evolution of the star. The He-shell values are enclosed in boxes, to help judge the match with SiC values.

The isotope ratios (on the right) provide a particularly stringent test and at least three agree very closely. $^{20}\text{Ne}/^{22}\text{Ne}$ is a little high, but only because it was not corrected for the N component. We do not show $^3\text{He}/^4\text{He}$ because it contains a third component, but the (fortuitously linear) trend on a plot against $^{20}\text{Ne}/^{22}\text{Ne}$ (Fig. 2*c*) suggests that $^3\text{He}/^4\text{He}$ in the G component is quite low, less than a third of the solar value of 1.42×10^{-4} . The elemental abundances are also quite similar, and show the hallmarks of AGB-star He shells: high ^{22}Ne , Kr-S and Xe-S, from He burning of ^{14}N and from the s-process, and low ^{36}Ar , because Ar is too heavy to be made by hydrostatic He burning and too light to be made by neutron capture on Fe,

but is actually depleted by neutron capture. The observed depletion is just of the right order; the ($^{22}\text{Ne}/^{36}\text{Ar}$)_G ratios of SiC and He shells both are $\sim 10^3$ times solar. Only ^{130}Xe is distinctly too high. This problem will be discussed below.

It is remarkable that not only isotope ratios but even elemental abundances (except for ^{130}Xe) agree so well with predictions for He shells. Apparently, at least the lighter noble gases were trapped in SiC without significant fractionation, that is, by a potentially non-selective process such as ion implantation from a stellar wind, and without subsequent diffusion loss of the lighter gases. Meteoritic SiC seems to contain some pristine AGB-star He-shell matter.

Origin of SiC and its noble gas components

In contrast to the G component, the N component looks fractionated relative to its source (assumed to be solar, to a first approximation). The pattern is flat from He to Ne, but then rises gently to Ar and more steeply to Kr and Xe (Fig. 3). The nearly flat trend for the three lighter gases again suggests ion implantation without fractionation. The strong enrichment of Kr_N and Xe_N, 10^2 – 10^4 times solar, cannot possibly be due to nuclear processes, but could arise in a stellar wind from a partially ionized region—

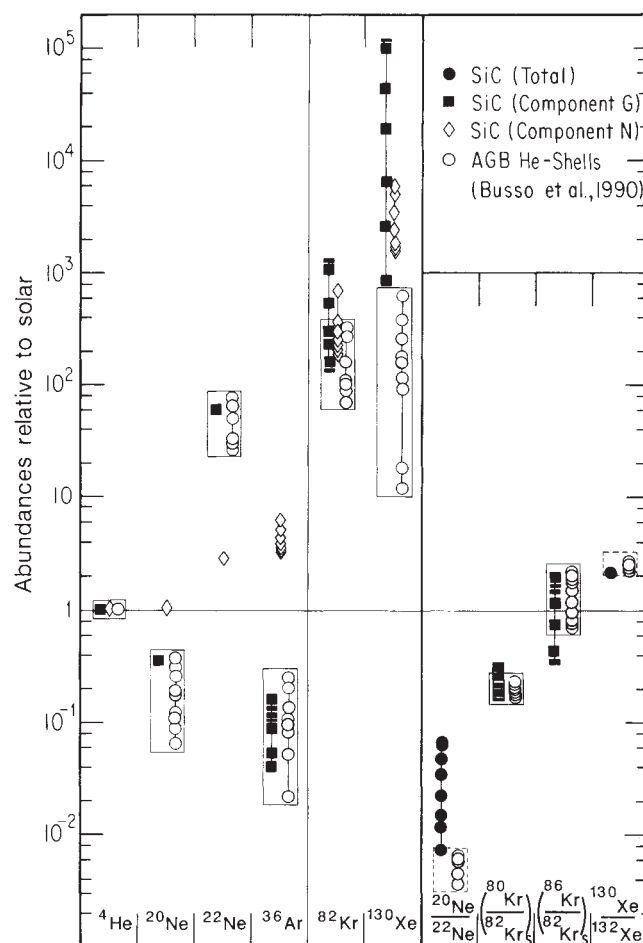


FIG. 3 Right, isotope ratios of SiC and AGB-star He shells, normalized to solar ratios. Boxes indicate known range of He-shell values^{13,14}. The sets are remarkably similar, even when the SiC values are not corrected for the N component (Ne, Xe). Left, Elemental abundances in SiC and He shells, normalized to solar values and ^4He . The SiC values have been resolved into isotopically anomalous (G) and normal (N) components. Abundances of the G component closely resemble those for AGB-star He shells, suggesting that the G component came from He shells and was trapped in SiC by a chemically non-selective process such as ion implantation, without subsequent diffusion loss of lighter gases.

the heavy gases, having lower ionization potentials, would be preferentially ionized and accelerated. (This mechanism was invoked²² to explain the ~1,800-fold enrichment of Ba-S over Xe-S).

Two important clues are: (1) the close coupling of Kr_N and Xe_N with Kr_G and Xe_G ; (2) the linear correlation of $^{86}Kr/^{82}Kr$ with Xe/Kr . The first trend is most clearly shown by the constancy of the isotope ratios $^{84}Kr/^{82}Kr$ and $^{130}Xe/^{132}Xe$, especially in stepped heating (Fig. 1a, b) but even in bulk samples (Table 1, Fig. 3). The second trend is illustrated in Fig. 4, for both the measured data and the G component alone. The inset in Fig. 4 gives another illustration of the constant N/G ratio: the lowest isotope ratios seen in stepped heating. Compared to the ratios for bulk samples (Table 1), they minimize contributions by loosely bound and hence probably extraneous Kr and Xe. The $^{136}Xe/^{130}Xe$ ratio tabulated here is the most sensitive measure of the N/G ratio, because ^{136}Xe is not made in the s-process and merely survives in small amounts owing to its low neutron capture cross-section.

The simplest interpretation of the linear trend in Fig. 4 is that the SiC fractions are mixtures of two components of fixed composition and different but overlapping size distributions: a fine-grained Xe-rich blend of $Kr_{G,N}$ and $Xe_{G,N}$ in the lower right and a coarse-grained Xe-poor blend in the upper left, of similar G/N ratio but higher $^{86}Kr/^{82}Kr$ (3.18 ± 0.06 for the pure G component). Alternatively, there could be a continuum of fortuitously collinear compositions.

The endmember in the upper left is not a problem, as it matches He-shell values for AGB stars¹⁴ of low metallicity, either asymptotic or late pulse. The endmember in the lower right matches $^{86}Kr/^{82}Kr$ ratios of high metallicity but has much too high a $^{130}Xe/^{82}Kr$ ratio, by factors of 10^2 – 10^3 . It seems necessary to invoke 'chemical' factors, such as partial ionization, to explain the high Xe/Kr ratios in Fig. 4. The following scenario seems to fit the data, at least qualitatively.

SiC grains condense from the expanding envelope of a pulsating AGB star at less than about two stellar radii²³ and move outward at 5 – 10 km s^{-1} , while becoming impregnated with ions from a stellar wind. Chemically active elements such as N, Al

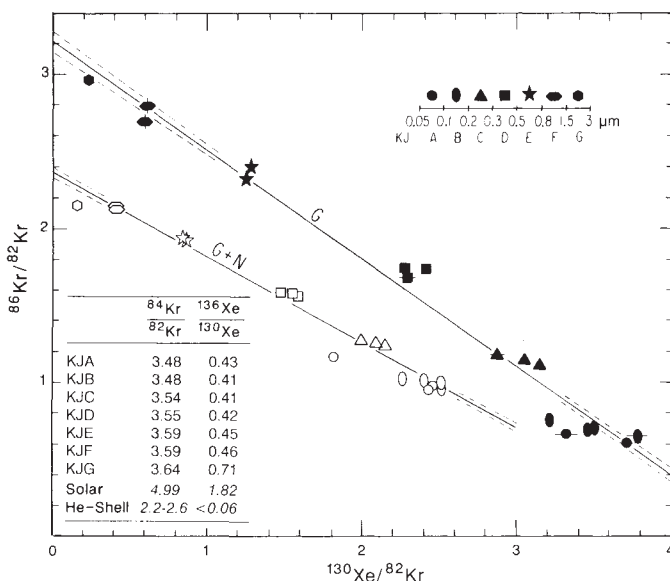


FIG. 4 $^{86}Kr/^{82}Kr$ plotted against $^{130}Xe/^{82}Kr$. The linear correlation suggests that the SiC size fractions are binary mixtures of a fine-grained Xe-rich blend of $Kr_{G,N}$ and $Xe_{G,N}$ in the lower right and a coarse-grained Xe-poor blend in the upper left, of similar G/N ratio but higher $^{86}Kr/^{82}Kr$. The Xe enrichment may reflect preferential acceleration of Xe in a stellar wind from a partially ionized region.

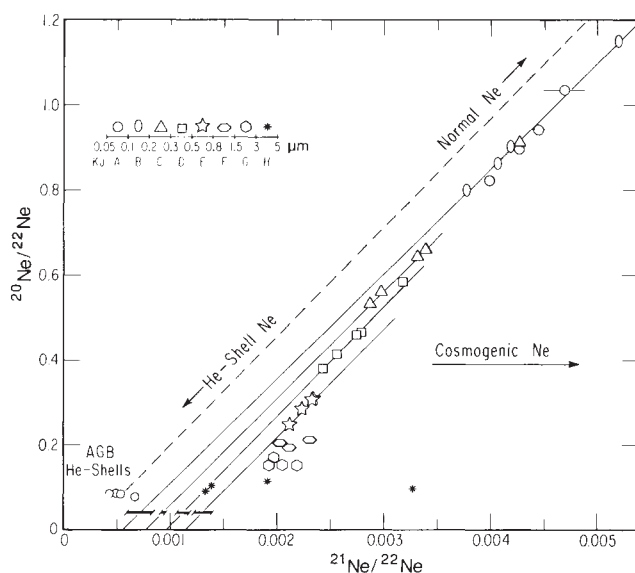


FIG. 5 $^{20}Ne/^{22}Ne$ plotted against $^{21}Ne/^{22}Ne$. With increasing grain size, samples lie farther to the right of the mixing line for a binary mixture of normal Ne (offscale to the upper right) and He-shell Ne (lower left), indicating progressively larger amounts of cosmogenic ^{21}Ne . $^{20}Ne/^{22}Ne$ ratios fall with increasing grain size, reflecting larger proportions of He-shell Ne. The four regression lines are for temperature fractions of samples KJB, C, D and E. The isotope ratios for He-shell Ne and normal Ne are $^{21}Ne/^{22}Ne = 5.4 \times 10^{-3}$ and 3.3×10^{-2} , $^{20}Ne/^{22}Ne = 0.081$ and 8.4 . The latter component approximates planetary and galactic Ne and lies on regression lines through the temperature fractions of the KJB, KJC and KJD. The He-shell mean¹⁴ is for $Z = 0.009$ – 0.02 and an overlap factor $r = 0.6$; the analogous mean for $r = 0.8$ lies slightly to the right of the KJB line and thus is too high, giving a physically meaningless, negative ^{21}Ne residual.

and perhaps Ba are taken up by SiC according to their condensation temperatures and affinities. Noble gases, on the other hand, are taken up according to the abundance of their ions in the wind, which reflects ionization and acceleration conditions. Two wind components (or a continuum) seem to be required. (1) A minor component from a hot, fully ionized region, which contributes most of the He, Ne and Ar in unfractionated ratios. The N-component for these three gases remains nearly constant with increasing grain size and $^{86}Kr/^{82}Kr$, but the G component increases (the He, Ne and Ar isotope ratios shift toward He-shell values; Table 1), as expected with increasing dredge-up of He-shell material. (2) A component from a cooler, partially ionized region, which is strongly enriched in Kr, Xe and perhaps Ba. During the early pulses, while $^{86}Kr/^{82}Kr$ is low, the temperature is low enough to give a high Xe/Kr ratio. Later, $^{86}Kr/^{82}Kr$ and the temperature rise, yielding progressively lower Xe/Kr ratios. The total fluence implied by the 4He contents is $\sim 10^8 \text{ g}^{-1}$ or $\sim 10^{15} \text{ cm}^{-2}$.

Equivalent results could be obtained from two stars: one with high metallicity, producing fine-grained SiC and a cool wind, and the other its opposite in all the respects. An alternative, more speculative scenario would be to vary the composition of the stellar wind with pulsation phase. Between pulses, the wind would consist of cool, only partially ionized envelope material, enriched in the heavier noble gases. If dredge-up after pulses does not yield complete mixing during ascent, then material in convection-cell hotspots could perhaps produce a second, unfractionated wind component of He-shell composition.

It remains to be seen whether either model can quantitatively account for the trends in Figs 3 and 4—especially the near constancy of the G/N ratio—either for single stars or for the average of many stars.

TABLE 2 Presolar cosmic-ray exposure ages of SiC

Sample	Size (μm)	^{22}Ne ($10^{-8} \text{ ml g}^{-1}$)	$^{21}\text{Ne}_c$	f_r^*	Age (Myr)
KJ	0.05–10	17,500	14.3	0.46	72 ± 11
KJA	0.05–0.1	5,100	2.8	0.23	27 ± 6
KJB	0.1–0.2	9,300	4.6	0.33	32 ± 8
KJC	0.2–0.3	13,400	8.1	0.41	46 ± 9
KJD	0.3–0.5	19,600	14.7	0.50	69 ± 11
KJE	0.5–0.8	28,400	26.7	0.59	106 ± 13
KJF	0.8–1.5	35,800	39.5	0.69	135 ± 15
KJG	1.5–3	28,800	35.4	0.79	106 ± 10
KJH	3–5	6,300	7.2	0.86	19 ± 2
LQB	0.05–0.15	2,060	1.1	0.28	7 ± 2
LQC	0.15–0.3	3,750	2.8	0.40	15 ± 3
LQD	0.3–0.5	6,330	6.3	0.50	29 ± 4
LQE	0.5–1	6,080	8.0	0.62	30 ± 3
LQF	1–2	3,880	7.0	0.75	22 ± 1

* Fraction of ^{21}Ne recoils retained^{26,31}.

Neon-E(H): not from ^{22}Na ?

Our data have important consequences for the origin of Ne-E(H), the high-temperature type^{24,25} of Ne-E, which is the anomalous neon component in SiC (Fig. 5). It is usually interpreted as a mixture of pure ^{22}Ne (from decay of ^{22}Na , ref. 10) with a normal 'planetary' component. When SiC samples became available for study, the data points turned out to lie consistently to the right of a mixing line between ^{22}Ne and planetary Ne, but this deviation has been attributed to a small spallogenic component rich in ^{21}Ne , produced in a presolar cosmic-ray irradiation of some 40 Myr^{5,9,26}.

Figure 5 shows that He-shell Ne lies to the left of the regression lines and below the lowest SiC point, and thus is no less plausible an endmember than is pure ^{22}Ne . Indeed, there is a strong reason for believing that it is a more plausible endmember. The match of the G-component elemental abundances in Fig. 3 suggests that the first four noble gases were trapped without fractionation, and because ^{22}Ne is present in He-shell proportions, there is no room for additional ^{22}Ne from ^{22}Na . Thus Ne-E(H) seems to be 'parentless' nucleosynthetic Ne from the He shell, diluted by variable amounts of Ne from the envelope. (A nucleosynthetic origin had been considered previously^{27,29}, but under most conditions examined, the predicted $^4\text{He}/^{22}\text{Ne}$ or $^{20}\text{Ne}/^{22}\text{Ne}$ ratios were too high.) The implantation process need not be efficient, as $^{22}\text{Ne}/\text{Si}$ is only 6.4×10^{-7} in the most enriched sample, compared to ~ 30 in an AGB-star He shell^{13,14}.

Age of SiC

Obviously, we need to reconsider the presolar cosmic-ray exposure age of SiC, nominally the interval between formation of the SiC grains and their arrival in the solar nebula. In contrast to previous calculations, we use He-shell Ne rather than pure ^{22}Ne as the Ne-E endmember. The samples again contain excess $^{21}\text{Ne}_c$: all lie to the right of a mixing line between mean He-shell Ne and representative normal Ne (Fig. 5). We have resolved excess ^{21}Ne contents relative to this mixing line, and have calculated nominal presolar exposure ages as before^{5,15,26}, using a ^{21}Ne production rate³⁰ of $0.421 \times 10^{-8} \text{ ml g}^{-1} \text{ Myr}^{-1}$, corrections for the recent cosmic-ray irradiation ($0.23 \times 10^{-8} \text{ ml g}^{-1}$) and recoil losses³¹ (Table 2).

The ages vary with grain size and range up to 135 Myr, well above our previous values of ~ 40 Myr. An argument for the validity of these ages is that the corresponding corrections for cosmogenic ^3He cause three-isotope plots involving $^3\text{He}/^4\text{He}$ to become linear and to extrapolate to $^3\text{He} = 0$ for the He shell. Surprisingly, the LQ samples have systematically lower ages and lower ^{22}Ne contents than their KJ counterparts, implying that both ^{21}Ne and ^{22}Ne were lost during the lengthier and more drastic chemical processing of the LQ samples. The lost gases

were probably located in gas-rich grains rather than in surface layers, otherwise Ne contents would fall rather than rise (Table 2) with increasing grain size.

Figure 6 shows that the ages and ^{22}Ne contents correlate rather well, as expected for a mixture of gas-rich, old, chemically reactive (perhaps radiation-damaged) grains with gas-poor, young, chemically resistant grains. The slope is +1, suggesting that the gas-rich grains have similar mean ages, regardless of size. (An analogous plot of ^{21}Ne contents rather than ^{21}Ne ages gives a poorer correlation and smaller slope; the key difference is that the ages include corrections for recoil losses of up to a factor of four). The gas-poor grains either formed from gas-rich grains in a late heating event (perhaps in the early Solar System) that outgassed and annealed them or, less likely, from one or more AGB stars that slightly predated the Solar System and had unusually weak stellar winds. Either way, the ages in Table 2 underestimate the age of the old grains, owing to dilution by young grains. The greatest ages in Table 2 are, however, only three to four times shorter than the predicted age of refractory interstellar grains, 400 Myr (ref. 32). We do not know the proportion of gas-rich grains in individual fractions, but at least for the parent sample, KJ, it seems to be less than about a third (judged from yields in different separation procedures), which would imply that the old grains in KJ are ≥ 200 Myr older than the Solar System.

Apparently the 40-Myr age of meteoritic SiC can now be attributed to one of the mechanisms originally proposed: degassing of much of the SiC just before or during the formation of the Solar System. Moreover, if the scarcity of SiC in the interstellar medium is due to a rapid destruction process³³, then this process must be sufficiently variable to permit some old SiC to survive.

Three main conclusions emerge from our work. (1) SiC can be sorted according to $^{86}\text{Kr}/^{82}\text{Kr}$ and hence stellar source by grain-size separation. (2) The elemental and isotopic pattern of noble gases in SiC is strikingly similar to that of low-mass AGB stars, showing that SiC contains a well preserved record of nuclear and chemical processes in its parent stars. Isotopes of

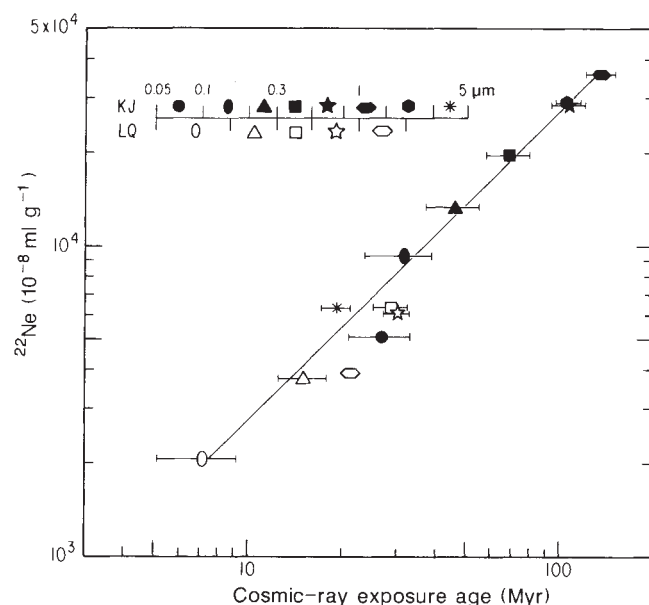


FIG. 6 ^{22}Ne content plotted against cosmic-ray exposure age; the slope of the line is one. Both ^{22}Ne content and ages are systematically lower for LQ samples, which underwent longer chemical treatments. The samples are apparently mixtures of gas-rich, old, chemically reactive (perhaps radiation damaged) grains and gas-poor, young, resistant grains. The latter may have been degassed and annealed during formation of the Solar System.

at least six other elements can be measured in SiC, and should provide still more detailed information. (3) A large part of the SiC was degassed during or just before formation of the Solar System, but a minor part has remained pristine. Although the

surviving SiC made only a very small contribution to the early Solar System (4×10^{-5} of the Si in primitive meteorites⁴), it is a new and potentially informative window on the Galaxy 4.6×10^9 yr ago. □

Received 24 July; accepted 3 October 1990.

1. Srinivasan, B. & Anders, E. *Science* **201**, 51–56 (1978).
2. Clayton, D. D. & Ward, R. A. *Astrophys. J.* **224**, 1000–1006 (1978).
3. Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **52**, 1235–1244 (1988).
4. Tang, M., Anders, E., Hoppe, P. & Zinner, E. *Nature* **339**, 351–354 (1989).
5. Zinner, E., Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **53**, 3273–3290 (1989).
6. Matsuda, J., Lewis, R. S. & Anders, E. *Astrophys. J.* **237**, L21–L23 (1980).
7. Ott, U., Begemann, F., Yang, J. & Epstein, S. *Nature* **332**, 700–702 (1988).
8. Lewis, R. S. & Amari, S. *Meteoritics* **24**, 293 (1989).
9. Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **52**, 1245–1254 (1988).
10. Clayton, D. D. *Nature* **257**, 36–37 (1975).
11. Wopenka, B. et al. *Meteoritics* **24**, 342 (1989).
12. Beer, H. & Macklin, R. L. *Astrophys. J.* **339**, 962–977 (1989).
13. Busso, M., Gallino, R., Picchio, G. & Raiteri, C. M. *Proc. Int. Symp. on Nuclear Astrophysics, Nuclei in the Cosmos* (eds Oberhummer, H. & Hillebrandt, W.) (Max Planck Institute for Physics and Astrophysics, Garching, in the press).
14. Gallino, R., Busso, M., Picchio, G. & Raiteri, C. M. *Nature* **348**, 298–302 (1990).
15. Amari, S. & Lewis, R. S. *Meteoritics* **24**, 247–248 (1989).
16. Tang, M. et al. *Geochim. cosmochim. Acta* **52**, 1221–1234 (1988).
17. York, D. *Can. J. Phys.* **44**, 1079–1086 (1966).

18. Amari, S., Anders, E., Virag, A. & Zinner, E. *Nature* **345**, 238–240 (1990).
19. Eugster, O., Eberhardt, P. & Geiss, J. *J. geophys. Res.* **74**, 3874–3896 (1969).
20. Gallino, R., Busso, M., Picchio, G., Raiteri, C. M. & Renzini, A. *Astrophys. J.* **334**, L45–L49 (1988).
21. Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–214 (1989).
22. Ott, U. & Begemann, F. *Astrophys. J.* **353**, L57–L60 (1990).
23. Draine, B. T. in *Physical Processes in Red Giants* (eds Iben, I. Jr & Renzini, A.) 317–333 (Reidel, Dordrecht, 1981).
24. Lewis, R. S., Alaerts, L., Matsuda, J. & Anders, E. *Astrophys. J.* **234**, L165–L168 (1979).
25. Eberhardt, P., Jungck, M. H. A., Meier, F. O. & Niederer, F. *Astrophys. J.* **234**, L169–L171 (1979).
26. Tang, M. & Anders, E. *Astrophys. J.* **335**, L31–L34 (1988).
27. Black, D. C. *Geochim. cosmochim. Acta* **36**, 377–394 (1972).
28. Arnould, M. & Nørgaard, H. *Astr. Astrophys.* **64**, 195–213 (1978).
29. Hillebrandt, W. & Thielemann, F.-K. *Astrophys. J.* **255**, 617–623 (1982).
30. Reedy, R. C. *Lunar planet. Sci.* **20**, 888–889 (1989).
31. Ray, J. & Völk, H. J. *Icarus* **54**, 406–416 (1983).
32. McKee, C. *IAU Symp.* **135**, 431–433 (1989).
33. Whittet, D. C. B., Duley, W. W. & Martin, P. G. *Mon. Not. R. astr. Soc.* **244**, 427–431 (1990).

ACKNOWLEDGEMENTS. We thank R. Gallino and colleagues for data and advice, particularly for suggesting that AGB stars may account for Ne-E, and D. D. Clayton and U. Ott for reviews. This work was supported in part by NASA.

On the astrophysical interpretation of isotope anomalies in meteoritic SiC grains

R. Gallino*, M. Busso†, G. Picchio† & C. M. Raiteri‡

* Istituto di Fisica generale dell' Università, Via P. Giuria 1, 10125 Torino, Italy

† Osservatorio Astronomico di Torino, Strada dell'Osservatorio 20, 10025 Pino Torinese, Italy

‡ International School for Advanced Studies, Via Beirut 4, 34014 Trieste, Italy

Meteoritic silicon carbide grains, formed in the winds from carbon stars, contain noble gases and other species whose elemental and isotopic abundances are a probe of stellar nucleosynthesis. Theoretical models of carbon stars can explain a variety of measured abundances, in particular the range of krypton isotope ratios and the excess ^{22}Ne found in the grains.

In the companion paper, Lewis *et al.*¹ present measurements of anomalous isotopic and elemental abundances of noble gases in SiC grains of extreme purity extracted from the Murchison meteorite. Among the carriers preserving the cosmic memory² of their nucleosynthetic origin, SiC grains in primitive meteorites are characterized by large anomalies in C, N, Si^{3–5} and especially noble gases^{3,6}, the isotope ratios of which are typical of slow neutron capture (the s-process), suggesting their formation in the extended atmospheres of carbon stars^{5,6}.

These results allow us to test models of nucleosynthesis in low-mass stars ($1 \leq M/M_{\odot} \leq 3$) during the thermally pulsing asymptotic giant branch (TP-AGB) phase. These stars become rich in C and s-process elements at their surface through mixing of material from the partially He-processed region⁷, and have been shown^{8,9} to reproduce the distribution of s-process isotopes heavier than atomic mass number $A \sim 80$ in the Solar System, the so-called main component.

The models

We have performed calculations for a series of nucleosynthesis models for low-mass TP-AGB stars of different metallicity Z , which is defined as the initial mass fraction of all nuclei heavier than He. We distinguish Z from the iron content, for which we shall adopt the common logarithmic notation of stellar spectroscopy: $[\text{Fe}/\text{H}] = \log(N_{\text{Fe}}/N_{\text{H}})_{\text{star}} - \log(N_{\text{Fe}}/N_{\text{H}})_{\odot}$, where $N_{\text{Fe}}/N_{\text{H}}$ is the atom ratio of iron to hydrogen.

The details of the models and a description of the large reaction scheme used, which is based on 450 nuclei from He to Po, will be presented elsewhere. Here we recall that during thermal pulses He is partially converted into C, whereas ^{14}N (which is formed in the previous processing of the original C, N and O nuclei during H burning) is transformed into ^{22}Ne through the chain: $^{14}\text{N}(\alpha, \gamma)^{18}\text{F}(\beta^+ \nu)^{18}\text{O}(\alpha, \gamma)^{22}\text{Ne}$. For the neutron source, we followed recent suggestions^{10–12} for the formation of a small amount of ^{13}C ($\sim 10^{-5} M_{\odot}$) in the intershell region; this is engulfed by the subsequent expansion of the convective He shell, where the reaction $^{13}\text{C}(\alpha, n)^{16}\text{O}$ easily occurs at moderate temperatures ($T \approx 1.5 \times 10^8$ K). The neutrons released are captured by the iron seeds and the lighter nuclei, such as Mg or Si isotopes, imprinting a distinctive s-process signature on almost all elements.

According to current models, activation of the ^{13}C neutron source is likely to occur in iron-poor stars belonging both to the halo of the Galaxy (the so-called population II, with $[\text{Fe}/\text{H}] \leq -1$) and to the galactic disk (old population I stars with $[\text{Fe}/\text{H}] \leq -0.5$). We have assumed that it is also effective for population I stars with higher iron content, up to solar, in the galactic disk. Arguments in favour of this have been recently presented^{8,13}, although they still need to be verified by stellar-evolution calculations (see ref. 14 for a discussion). This hypothesis is strengthened *a posteriori* by the fact that it allows good fits to the abundance distributions of stars of spectral type MS, S and

TABLE 1 Expected isotope ratios from AGB stars

Ratios	He shell, asymptotic, $r=0.6$			Envelope, $C/O=1$, $r=0.6$			Measured ¹ in SiC grains
	[Fe/H] = -0.6 $\tau_0=0.32$	[Fe/H] = -0.3 $\tau_0=0.22$	[Fe/H] = 0.0 $\tau_0=0.12$	[Fe/H] = -0.6	[Fe/H] = -0.3	[Fe/H] = 0.0	
$^{12}\text{C}/^{13}\text{C}$	∞	∞	∞	78.0	40.2	40.3	$41 \pm 3^*$
$^3\text{He}/^4\text{He}$	0	0	0	9.58×10^{-4}	9.63×10^{-4}	9.09×10^{-4}	$(0.5-2) \times 10^{-4}$
$^4\text{He}/^{22}\text{Ne}$	4.22×10^2	3.84×10^2	1.79×10^2	1.20×10^4	1.06×10^4	3.00×10^3	$(2.4-9.4) \times 10^2$
$^{20}\text{Ne}/^{22}\text{Ne}$	4.91×10^{-2}	8.08×10^{-2}	8.45×10^{-2}	3.60	5.75	3.44	0.098-0.91
$^{21}\text{Ne}/^{22}\text{Ne}$	5.58×10^{-4}	6.85×10^{-4}	4.96×10^{-4}	9.05×10^{-3}	1.42×10^{-2}	8.57×10^{-3}	$(1.7-4.3) \times 10^{-3}$
$^{36}\text{Ar}/^{22}\text{Ne}$	6.15×10^{-4}	1.14×10^{-3}	1.42×10^{-3}	9.13×10^{-2}	1.52×10^{-1}	8.97×10^{-2}	0.0129-0.110
$^{38}\text{Ar}/^{36}\text{Ar}$	0.84	0.75	0.57	0.19	0.19	0.19	0.19-0.23
$^{83}\text{Kr}/^{82}\text{Kr}$	0.35	0.35	0.34	0.42	0.48	0.53	0.62-0.72
$^{84}\text{Kr}/^{82}\text{Kr}$	2.39	2.31	2.07	2.62	2.77	2.84	3.6-4.5
$^{86}\text{Kr}/^{82}\text{Kr}$	3.22	2.27	1.08	2.78	1.80	1.11	1.01-2.06
$^{130}\text{Xe}/^{132}\text{Xe}$	0.39	0.40	0.43	0.36	0.26	0.19	0.16-0.36

We have assumed a ratio of $^{12}\text{C}/^{13}\text{C}=12$ in the envelope after the first dredge-up. The mass of the model star is $1.5 M_{\odot}$.

* Mass-averaged mean for another set of Murchison SiC (ref. 6).

C (subtype N) in the normal disk population^{7,13}.

In all calculations, we allowed the same amount of ^{13}C to burn but used different values of the overlap factor r , the amount of material previously processed in the He shell that remains for the next pulse. In a series of convective thermal instabilities, each characterized by a neutron exposure time $\Delta\tau = \int N_n(t')v(t')dt'$ (where N_n is the neutron density, $v = (2kT/\mu)^{1/2}$ is the thermal velocity and μ is the reduced mass), the overlap factor determines the mean exposure τ_0 of the seeds to asymptotic conditions, through the relation¹⁵ $\tau_0 = -\Delta\tau/\ln r$. The value of r in disk stars is uncertain. Recent calculations suggest that the interflash periods in TP-AGB stars of population I have only half the duration of population II^{16,17}, implying that it should increase with the iron content, from $r=0.6$, the value typical of population II stars¹⁸, perhaps to $r \approx 0.8$. By allowing it to vary in that range, different neutron exposures can be obtained even in models with the same iron content (a similar effect would result from varying the amount of ^{13}C ingested, within a narrow range, from star to star⁸). We followed the nucleosynthesis for 20 thermal instabilities, asymptotic conditions being reached typically after 10-20 pulses. To find the enrichment in the stellar atmosphere, where SiC grains form, the abundance distribution in the C-rich region was mixed into the envelope assuming simple criteria for mass loss and dredge-up¹³. We did not consider the possible effects of hot H-burning cycles in the H shell. The composition of the envelope at the beginning of TP-AGB phase was based on observations of normal red giants, in particular for C and N isotopes^{7,19}.

Comparison with SiC measurements

Selected results from our calculations of the isotope ratios measured in meteoritic SiC are shown in Table 1 for cases with different iron content. The first model ([Fe/H] = -0.6) was run assuming an initial doubled O/Fe ratio, typical of the old disk population²⁰. Columns 2 to 4 of Table 1 refer to asymptotic conditions achieved in the convective He shell, and columns 5 to 7 show the composition of the envelope when the dredge-up causes the formation of a carbon star with C/O=1 (these abundances do not differ much from those at the nominal threshold for SiC formation²). The column headings show that the mean neutron exposure decreases with increasing iron content. A suggestive result that emerges from this comparison is that the ratios measured in SiC samples (column 8) almost always lie between those found in the He shell and those found in the envelope when C/O=1. The grains seem to carry the carbon isotopic composition of the envelope at the stage when a carbon star forms; the noble gases also generally show ratios similar to those calculated in the envelope at this stage, but contaminated by a purer shell material (especially $^4\text{He}/^{22}\text{Ne}$ and $^{20}\text{Ne}/^{22}\text{Ne}$). Exceptions for Kr isotopes are probably due to admixtures of normal Kr (see ref. 1).

We can analyse the results in more detail by looking at the first four isotope ratios listed in Table 1, which bear information about the various mixing phases in red giants, especially the first dredge-up, in which the convective envelope penetrates the H-exhausted region, and the third dredge-up, in which the products of He-shell burning are convected to the surface. Indeed, $^3\text{He}/^4\text{He}$ depends on the first dredge-up (which increases ^3He in the envelope¹⁹, whereas ^4He is only marginally affected); $^4\text{He}/^{22}\text{Ne}$ and $^{20}\text{Ne}/^{22}\text{Ne}$ are signatures of the third dredge-up (because ^{22}Ne is increased, whereas ^{20}Ne is nearly unchanged). Finally, $^{12}\text{C}/^{13}\text{C}$ depends on both, because the first dredge-up increases ^{13}C and reduces ^{12}C , whereas the third dredge-up enriches the envelope in ^{12}C (ref. 21).

From the ratio of $^{12}\text{C}/^{13}\text{C}$ in the envelope at C/O=1, we can discard models of low iron content with enhanced O/Fe, because they cannot reproduce the measured values (see Table 1, column 2). This is supported by the analysis of the trend of $^4\text{He}/^{22}\text{Ne}$ with $^{20}\text{Ne}/^{22}\text{Ne}$ in the SiC samples. The linear correlation found in ref. 1 (see their Fig. 2) suggests a mixture of only two components, one of pure He-shell composition (their G component), the other made of envelope material at the moment of the formation of SiC grains (their N component). As Fig. 1 shows, the points defined by the shell material for models with an initially solar composition are close to the regression line through the measurements. It is possible, however, that the

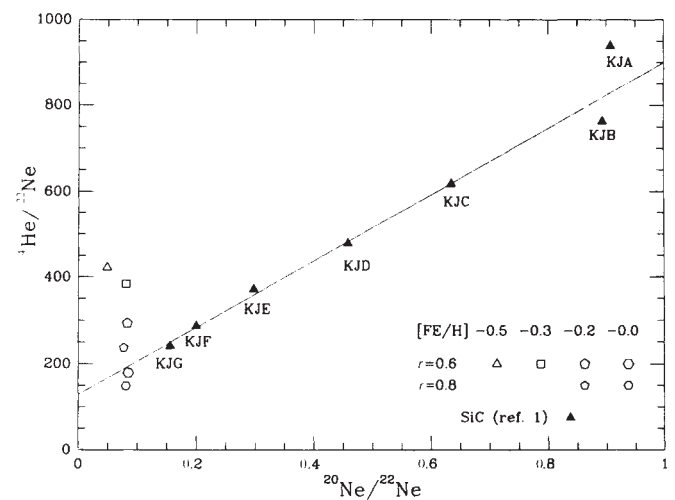


FIG. 1 $^4\text{He}/^{22}\text{Ne}$ plotted against $^{20}\text{Ne}/^{22}\text{Ne}$ for the measured ratios for Murchison SiC grains¹ (filled symbols) and the calculated values (open symbols).

TABLE 2 Expected isotope ratios for noble gases from population I AGB stars

Ratios	He shell, asymptotic		He shell, asymptotic		G component ¹
	[Fe/H] = -0.2 $r=0.6$, $\tau_0=0.17$	[Fe/H] = -0.2 $r=0.8$, $\tau_0=0.32$	[Fe/H] = 0.0 $r=0.6$, $\tau_0=0.12$	[Fe/H] = 0.0 $r=0.8$, $\tau_0=0.22$	
$^4\text{He}/^{22}\text{Ne}$	2.93×10^2	2.37×10^2	1.79×10^2	1.49×10^2	~ 180
$^{20}\text{Ne}/^{22}\text{Ne}$	8.25×10^{-2}	7.64×10^{-2}	8.45×10^{-2}	8.10×10^{-2}	$8.1 \times 10^{-2*}$
$^{21}\text{Ne}/^{22}\text{Ne}$	6.07×10^{-4}	9.62×10^{-4}	4.96×10^{-4}	7.78×10^{-4}	$5.4 \times 10^{-3*}$
$^{36}\text{Ar}/^{22}\text{Ne}$	1.24×10^{-3}	7.48×10^{-4}	1.42×10^{-3}	9.42×10^{-4}	$6.7 (-4)$
$^{38}\text{Ar}/^{36}\text{Ar}$	0.69	1.25	0.57	1.03	0.74^*
$^{80}\text{Kr}/^{82}\text{Kr}$	3.81×10^{-2}	3.21×10^{-2}	4.45×10^{-2}	3.59×10^{-2}	$(2-5) \times 10^{-2}$
$^{83}\text{Kr}/^{82}\text{Kr}$	0.35	0.35	0.34	0.35	0.329
$^{84}\text{Kr}/^{82}\text{Kr}$	2.21	2.44	2.07	2.42	2.55^*
$^{86}\text{Kr}/^{82}\text{Kr}$	1.78	2.62	1.08	1.79	$0.57-2.96$
$^{130}\text{Xe}/^{132}\text{Xe}$	0.41	0.39	0.43	0.40	0.49

* Adopted as typical from AGB models.

was slightly altered by fractionation in the stellar wind; using a fractionation factor 0.54 deduced from the comparison of the solar wind and the photosphere²², the best-fit model is one with an iron content slightly lower than solar, $[\text{Fe}/\text{H}] \approx -0.2$, but still a population I model.

A special comment is required on the ratio $^{38}\text{Ar}/^{36}\text{Ar}$ in Table 1: at first sight, it may seem strange that the values estimated for the envelope and the SiC measurements are always far from He-shell conditions but close to solar. This is because both ^{36}Ar and ^{38}Ar are only slightly affected by s-processing, being destroyed and produced by neutron captures, respectively, by a factor of only about two. Consequently, even mixing a huge amount of shell material into the envelope does not change their ratio much.

From the comparisons discussed so far, two main conclusions can be drawn: (1) the excess ^{22}Ne measured in the SiC samples (the high-temperature Ne component Ne-E(H) discovered in primitive chondrites^{23,24}) comes from the ^{14}N chain of reactions in the He-shell zone, without the need of the generally accepted radiogenic contribution from ^{22}Na (see, for example, ref. 2); (2) the isotope anomalies shown by SiC grains are reproduced by carbon stars of population I, with metallicity Z in the range 0.01–0.02. For such stars, Table 2 gives the noble gas composition in the He shell (using an overlap factor r of 0.6 or 0.8 to account for its uncertainty in population I red giants), together with the expectations for the G component¹.

For all cases in Table 2 $^4\text{He}/^{22}\text{Ne}$ and $^{20}\text{Ne}/^{22}\text{Ne}$ are in good agreement with SiC extrapolations, if one allows a slight He/Ne fractionation. For $^{21}\text{Ne}/^{22}\text{Ne}$, we note that there is a small nucleosynthetic production of ^{21}Ne , which is essentially determined by neutron captures of ^{20}Ne . The enhancement in the He shell over the initial abundance does not exceed a factor of five, but it is important to take it into account, owing to a cosmogenic contribution to the abundance of this isotope, from which a cosmic-ray exposure age of the SiC grains may be inferred¹. Good agreement is found for $^{36}\text{Ar}/^{22}\text{Ne}$, suggesting that this ratio is not significantly altered by fractionation, as is indeed observed in the solar wind²². Owing to their lower ionization potential, however, Kr and Xe may be fractionated with respect to Ne.

Finally, although the Kr isotopes deserve special discussion (below), the calculated $^{130}\text{Xe}/^{132}\text{Xe}$ ratio is about constant, and rather close to the value extrapolated in ref. 1, for all exposures considered; the differences can be understood in the light of the large uncertainties in the neutron capture cross-sections of Xe isotopes.

Constraints from Kr isotopes

The isotope anomalies of Kr in SiC grains reflect a complex situation. In particular, ^{80}Kr and ^{86}Kr lie at branch points on the neutron flow and thus are sensitive to temperature and neutron density; the signature of this might be found in

meteoritic anomalies²⁵. Before comparing our theoretical expectations with the measurements, we note that neutron capture cross-sections of all Kr isotopes are determined within 6–12% at 30 keV, and that some of them do not follow the usual $1/v$ law, adding a further source of uncertainty. This prevents us from deriving too stringent constraints. In this respect, ^{84}Kr is worth a special comment, because in models of TP-AGB stars the reproduction of the Solar System abundance of ^{86}Kr favours⁹ a choice for the ^{84}Kr cross-section near the lower limit of the present uncertainty range (34 ± 4 mbarn; H. Beer, personal communication). The results of Table 1 and 2 have been obtained using $\sigma(^{84}\text{Kr}) = 30$ mbarn and standard^{26,27} recommendations (see also ref. 9) for the other Kr isotopes, including a correction factor for the temperature-dependence of $\sigma(^{83}\text{Kr})$ ^{28,29}. We have also made some tests in which we varied the cross-sections within their present uncertainties; we found that the predicted ratios $^{83}\text{Kr}/^{82}\text{Kr}$ and $^{84}\text{Kr}/^{82}\text{Kr}$ lie in the range 0.28–0.40 and 2.2–2.6, respectively.

The most remarkable feature shown by Kr isotopes¹ is the large spread extrapolated for $^{86}\text{Kr}/^{82}\text{Kr}$ and the smaller one for $^{80}\text{Kr}/^{82}\text{Kr}$. Both can in principle be accounted for by two rather different scenarios involving either many stars or a single star. A first possibility is that these spreads come from different stars

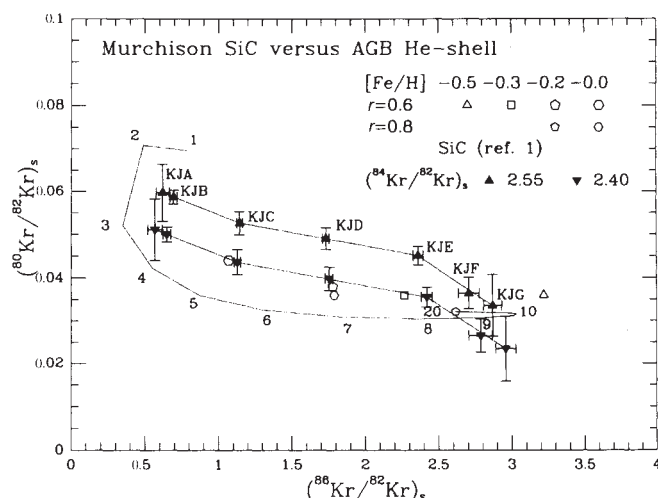


FIG. 2 $(^{80}\text{Kr}/^{82}\text{Kr})_s$ plotted against $(^{86}\text{Kr}/^{82}\text{Kr})_s$. The SiC extrapolated data (KJA to KJG) are scaled according to the two different choices for the ratio $^{84}\text{Kr}/^{82}\text{Kr}$ 2.55 (filled upright triangles) and 2.40 (filled inverted triangles). The second set of SiC values, in particular, agrees remarkably well with the calculated asymptotic He-shell expectations. These last are shown as open symbols, for the values of metallicity and of overlap factor r indicated. The variation of these values with the number of thermal pulses is shown, as an example, for the case $[\text{Fe}/\text{H}] = -0.2$, $r = 0.8$.

with a wide range of neutron exposures, up to $\tau \approx 0.30 \text{ mbarn}^{-1}$. Table 2 shows that both ratios do vary with the neutron exposure, the first one decreasing by 20%, the second one strongly increasing (by 50%). Figure 2 shows that the agreement between theoretical expectations and SiC extrapolations is satisfactory, despite the uncertainties in the cross-section; the agreement is very good when we assume a value of 2.40 for the G component of $^{84}\text{Kr}/^{82}\text{Kr}$ (slightly less than the compromise adopted by Lewis *et al.*¹). In this case, the constant theoretical estimate for $^{83}\text{Kr}/^{82}\text{Kr}$ is slightly reduced, and thus falls at the lower limits of SiC extrapolations.

We note that some discrepancy remains for $^{84}\text{Kr}/^{82}\text{Kr}$; also this ratio varies, by up to 10%, in models with the same overlap factor and different neutron exposure (see Table 2), whereas the SiC data seem to fall on a single correlation line, implying a single He-shell endmember rather than a spread. A small departure from linearity is not excluded, however, owing to the crowding of the measurements in a small range; a better evaluation of the neutron capture cross-section of ^{84}Kr (both to the ground and to the excited level of ^{85}Kr) is certainly desirable.

A second possibility is that SiC grains reflect the conditions in one star, at different times in its AGB history. This is illustrated in Fig. 3, where the evolution of Kr isotope ratios in the He shell with pulse number mimics their behaviour as a function of τ . This analysis cannot be pursued further because the details of the mixing of s-process material with the envelope are not yet sufficiently known. For a single star to be the source of all the anomalies, the conditions for the formation of SiC grains must occur early enough in the TP-AGB phase to preserve the signature of various s-processed episodes, but avoid the very first pulses where the $^{84}\text{Kr}/^{82}\text{Kr}$ ratio is too low. Such a possibility is compatible with present results for the dredge-up mechanism³⁰. In any case, to reproduce the whole measured range by a single star mean asymptotic exposure should be high ($\tau \approx 0.30$), not very different from that characterizing the Solar System main component and the S stars in the Galaxy^{8,9,13}.

Although Fig. 2 shows better agreement for the possibility of contributions from several stars, we cannot decide firmly between the two models when the uncertainties on the cross-section are taken into account. The age determinations in ref. 1 and their variation with grain size do, however, favour the multistar scenario. In both cases only a small percentage of SiC grains would experience the maximum neutron exposure, so that our findings are not in disagreement with the average value $\tau_0 \approx 0.17 \text{ mbarn}^{-1}$ inferred from measurements of Ba isotopes³¹ in bulk Murchison SiC.

Despite many uncertainties, the Kr measurements in SiC grains give us some important information. First, the low $^{80}\text{Kr}/^{82}\text{Kr}$ ratio is a clear indication of a low temperature in the He shell, which is characteristic of the ^{13}C source but not of the ^{22}Ne one⁹. Second, the large variation of $^{86}\text{Kr}/^{82}\text{Kr}$ suggests a range of neutron exposures in the He-shell s-processed zone³². Although both these features were anticipated on the basis of less precise measurements³³, made on acid-resistant residues from the Murchison meteorite containing a small percentage of SiC, the new data define the trends far more accurately and for all five noble gases. These data have made it possible to compare the observations and theory in a deeper way.

Correlations of Xe, Kr and He

In the data of ref. 1 there is a wide spread in the Xe and Kr elemental abundances and the values are very high compared to He. Our s-process expectations can approach only the lower limits of measured ranges; as an example, even in the case with the highest τ value we find $(^{130}\text{Xe}/^4\text{He})_s = 4.75 \times 10^{-8}$ and $(^{85}\text{Kr}/^4\text{He})_s = 5.2 \times 10^{-7}$, compared to extrapolations¹ of $(6 \times 10^{-8} - 8 \times 10^{-6})$ and $(2.8 \times 10^{-7} - 2.2 \times 10^{-6})$, respectively. Moreover, an anticorrelation exists between the Xe-s abundance and $(^{85}\text{Kr}/^{82}\text{Kr})_s$, Xe being very high in the finest grain sizes with low $^{86}\text{Kr}/^{82}\text{Kr}$ values. We think that, on nuclear grounds

TABLE 3 Expected isotope ratios for other elements

$^{25}\text{Mg}/^{24}\text{Mg}$	2.7	$^{26}\text{Mg}/^{24}\text{Mg}$	3.8	$^{28}\text{Si}/^{28}\text{Si}$	1.8
$^{30}\text{Si}/^{28}\text{Si}$	2.9	$^{33}\text{S}/^{32}\text{S}$	2.1	$^{34}\text{S}/^{32}\text{S}$	1.8
$^{36}\text{S}/^{32}\text{S}$	120	$^{40}\text{K}/^{39}\text{K}$	37	$^{41}\text{K}/^{39}\text{K}$	2.9
$^{42}\text{Ca}/^{40}\text{Ca}$	5.5	$^{43}\text{Ca}/^{40}\text{Ca}$	7.5	$^{44}\text{Ca}/^{40}\text{Ca}$	2.6
$^{46}\text{Ca}/^{40}\text{Ca}$	20	$^{48}\text{Ca}/^{40}\text{Ca}$	1.5	$^{46}\text{Ti}/^{48}\text{Ti}$	4.3
$^{47}\text{Ti}/^{48}\text{Ti}$	2.3	$^{49}\text{Ti}/^{48}\text{Ti}$	7.9	$^{50}\text{Ti}/^{48}\text{Ti}$	20
$^{53}\text{Cr}/^{52}\text{Cr}$	1.3	$^{54}\text{Cr}/^{52}\text{Cr}$	17	$^{54}\text{Fe}/^{56}\text{Fe}$	67
$^{57}\text{Fe}/^{56}\text{Fe}$	4.2	$^{58}\text{Fe}/^{56}\text{Fe}$	8.0	$^{60}\text{Ni}/^{58}\text{Ni}$	8.9
$^{61}\text{Ni}/^{58}\text{Ni}$	70	$^{62}\text{Ni}/^{58}\text{Ni}$	46	$^{64}\text{Ni}/^{58}\text{Ni}$	280
$^{66}\text{Zn}/^{64}\text{Zn}$	14	$^{67}\text{Zn}/^{64}\text{Zn}$	21	$^{68}\text{Zn}/^{64}\text{Zn}$	30
$^{70}\text{Zn}/^{64}\text{Zn}$	2.7				
$^{28}\text{Si}/^{28}\text{Si}_0$	0.53	$^4\text{He}/^{28}\text{Si}$	4.7	$^{12}\text{C}/^{28}\text{Si}$	20
$^{24}\text{Mg}/^{28}\text{Si}$	1.1	$^{32}\text{S}/^{28}\text{Si}$	0.80	$^{39}\text{K}/^{28}\text{Si}$	2.0
$^{40}\text{Ca}/^{28}\text{Si}$	0.67	$^{48}\text{Ti}/^{28}\text{Si}$	0.64	$^{52}\text{Cr}/^{28}\text{Si}$	0.64
$^{56}\text{Fe}/^{28}\text{Si}$	0.64	$^{58}\text{Ni}/^{28}\text{Si}$	0.42	$^{64}\text{Zn}/^{28}\text{Si}$	3.0

Model AGB stars have $[\text{Fe}/\text{H}] = -0.2$, $\tau_0 = 0.17$ and $r = 0.6$. The lower set of values gives ratios of elements to ^{28}Si . To renormalize to other reference nuclei, ratios of ^{28}Si to ^4He and to ^{12}C are also indicated. All abundances are normalized to the solar ones, taken from ref. 22.

alone, there is no reasonable explanation for the high Xe and Kr values and for their anticorrelation. On the contrary, at the low exposures where $^{86}\text{Kr}/^{82}\text{Kr}$ is low, the expected nucleosynthetic Xe/Kr is the lowest. The above trends can, however, be explained by strong fractionation for both Kr and Xe, as suggested by Lewis *et al.*¹. In an ionized wind from a well mixed carbon-star envelope, such a fractionation would affect all the Kr and Xe nuclei, regardless of their origin, and is therefore expected to preserve the remarkable constancy of Xe isotope ratios found in the N component and in the G component¹. An elemental fractionation has recently been invoked to explain the high Ba/Xe ratio in the bulk Murchison SiC³².

Conclusions

Despite many unsolved problems, we have found close quantitative agreement between theoretical predictions of isotope abundances from the thermal pulsing in AGB stars and the abundances found in the SiC of the Murchison meteorite. We conclude that: (1) only the ^{13}C source acting at the relatively moderate temperatures ($T \approx 1.5 \times 10^8 \text{ K}$) of a thermal instability in low-mass TP-AGB stars reproduces the low $^{80}\text{Kr}/^{82}\text{Kr}$ ratios extrapolated

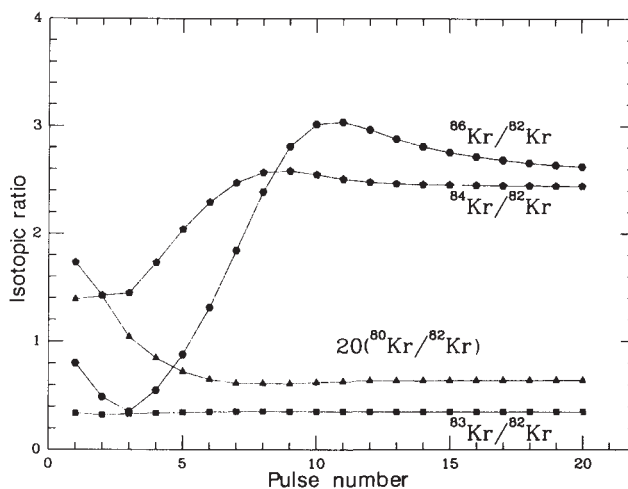


FIG. 3 Variations with pulse number of krypton isotope ratios with respect to ^{82}Kr for a TP-AGB star with $[\text{Fe}/\text{H}] = -0.2$. This model has an asymptotic neutron exposure $\tau_0 = 0.32 \text{ mbarn}^{-1}$ and a neutron exposure per pulse $\Delta\tau = 0.043 \text{ mbarn}^{-1}$. The values between pulses 4 and 10 cover almost the entire observed range for SiC.

for the pure s-component. This model also seems to be able to explain most of the other observed s-process signatures and the correlated He, Ne and Ar abundances. (2) The ^{13}C source has to operate in population I red giants with metallicity $Z = 0.01\text{--}0.02$. (3) The Ne–E(H) carried by SiC grains is explained by the excess of ^{22}Ne in the partially He-burned region. (4) The large spread of $^{86}\text{Kr}/^{82}\text{Kr}$ can be accounted for either by contributions from several giants with different neutron exposures or by different pulses in a single star.

We propose the following scenario. SiC grains form as thermal condensates in the cool mass-losing atmospheres of carbon stars, and move away into the circumstellar envelope. The noble gases are then implanted by stellar winds as ions; Xe and Kr, having lower ionization potentials, are more readily ionized and accelerated, whereas He and Ne require more extreme conditions. This

suggests winds with different degrees of ionization¹, either in a single star or in a multiplicity of stars.

As well as the noble gas and carbon isotope ratios discussed above, we can use TP–AGB models to predict a number of other anomalies for elements trapped in SiC grains. Table 3 gives our predictions for the He-shell abundances of selected nuclei up to Zn, using neutron capture cross-sections from ref. 26. We note, however, that these cross-sections need to be better determined, especially where poorly known (n, α) channels compete with (n, γ) ones, as is the case for the Si and S isotopes. Although the data in Table 3 are for a model with $\tau_0 = 0.17$, about midway in the range deduced from Kr isotopes, they do not depend much on this particular choice. Predictions for heavier elements would require a more detailed discussion because they are strongly related to the neutron exposure. □

Received 24 July; accepted October 1990.

- Lewis, R., Amari, S. & Anders, E. *Nature* **348**, 293–298 (1990).
- Clayton, D. D. *Q. J. R. astr. Soc.* **23**, 174–212 (1982).
- Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **52**, 1235–1244 (1988).
- Zinner, E., Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **53**, 3273–3290 (1989).
- Ott, U. in *Protostars and Planets* Vol. 3 (eds Levy, E. H. & Lunine, J. I.) (University of Arizona Press, in the press).
- Zinner, E., Tang, M. & Anders, E. *Nature* **330**, 730–732 (1987).
- Lambert, D. L. in *IAU Colloq. 106, Evolution of Peculiar Red Giant Stars* (eds Johnson, H. R. & Zuckerman, B.) 101–130 (Cambridge University Press, 1989).
- Gallino, R., Busso, M., Picchio, G., Raiteri, C. M. & Renzini, A. *Astrophys. J.* **334**, L45–L49 (1988).
- Käppeler, F., Gallino, R., Busso, M., Picchio, G. & Raiteri, C. M. *Astrophys. J.* **354**, 630–643 (1990).
- Iben, I. Jr & Renzini, A. *Astrophys. J.* **259**, L79–L83 (1982).
- Iben, I. Jr & Renzini, A. *Astrophys. J.* **263**, L23–L27 (1982).
- Iben, I. Jr & Renzini, A. *Astrophys. J.* **340**, 966–984 (1989).
- Busso, M., Gallino, R., Picchio, G. & Raiteri, C. M. *Proc. Int. Symp. Nuclear Astrophysics, Nuclei in the Cosmos* (eds Oberhummer, H. & Hillebrandt, W.) (Max Planck Institute for Physics and Astrophysics, Garching, in the press).
- Lattanzio, J. C. in *IAU Colloq. 106, Evolution of Peculiar Red Giant Stars* (eds Johnson, H. R. & Zuckerman, B.) 161–175 (Cambridge University Press, 1989).
- Ulrich, R. K. in *Explosive Nucleosynthesis* (eds Schramm, D. N. & Arnett, D. W.) 139–165 (University of Texas Press, Austin, 1973).
- Boothroyd, A. I. & Sackmann, I. J. *Astrophys. J.* **328**, 653–670 (1988).
- Lattanzio, J. C. *Astrophys. J.* **311**, 708–730 (1986).
- Iben, I. Jr *Astrophys. J.* **260**, 821–837 (1982).
- Iben, I. Jr in *7th Adv. Course Swiss Soc. Astr. Astrophys., Advanced Stages in Stellar Evolution* (eds Bouvier, P. & Maeder, A.) 3–146 (Geneva Observatory Press, 1977).
- Pagel, B. E. J. *Proc. Int. Symp. Nuclear Astrophysics, Nuclei in the Cosmos* (eds Oberhummer, H. & Hillebrandt, W.) (Max Planck Institute for Physics and Astrophysics, Garching, in the press).
- Iben, I. Jr & Renzini, A. *A. Rev. Astr. Astrophys.* **21**, 271–342 (1983).
- Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–214 (1989).
- Eberhardt, P., Jungck, M. H. A., Meier, F. O. & Niederer, F. *Astrophys. J.* **234**, L169–L171 (1979).
- Lewis, R. S., Alaerts, L., Matsuda, J. & Anders, E. *Astrophys. J.* **234**, L165–L168 (1979).
- Clayton, D. D. *Nature* **332**, 683–684 (1988).
- Bao, Z. Y. & Käppeler, F. *Atom Data nucl. Data Tables* **36**, 411–451 (1987).
- Beer, H. & Macklin, R. L. *Astrophys. J.* **339**, 962–977 (1989).
- Holmes, J. A., Woosley, S. E., Fowler, W. A. & Zimmermann, B. A. *Atom Data nucl. Data Tables* **18**, 305–412 (1976).
- Harris, M. J. *Astrophys. Space Sci.* **77**, 357–367 (1981).
- Lattanzio, J. C. *Astrophys. J.* **313**, L15–L18 (1987).
- Ott, U. & Begemann, F. *Astrophys. J.* **353**, L57–L60 (1990).
- Gallino, R., Busso, M., Picchio, G. & Raiteri, C. M. in *From Miras to Planetary Nebulae* (eds Mennessier, M. O. & Omont, A.) 329–331 (Editions Frontières, Paris, 1990).
- Ott, U., Begemann, F., Yang, J. & Epstein, S. *Nature* **332**, 700–702 (1988).

ACKNOWLEDGEMENTS. We thank E. Anders and his colleagues for support, H. Beer for discussions, and D. D. Clayton and U. Ott for many comments on the manuscript. Computer work was done at the Osservatorio Astronomico di Torino, the Istituto di Fisica Generale dell'Università di Torino and the Istituto Nazionale di Fisica Nucleare (Sezione di Torino).

The molecular mechanism by which insulin stimulates glycogen synthesis in mammalian skeletal muscle

Paul Dent, Alain Lavoine, Sara Nakielny, F. Barry Caudwell, Peter Watt* & Philip Cohen

Departments of Biochemistry and * Anatomy and Physiology, University of Dundee, Scotland, UK

The ability of insulin to promote the phosphorylation of some proteins and the dephosphorylation of others is paradoxical. An insulin-stimulated protein kinase is shown to activate the type-1 protein phosphatase that controls glycogen metabolism, by phosphorylating its regulatory subunit at a specific serine. Furthermore, the phosphorylation of this residue is stimulated by insulin *in vivo*. Increased and decreased phosphorylation of proteins by insulin can therefore be explained through the same basic underlying mechanism.

activates the protein tyrosine kinase associated with the receptor's β subunit¹. The next event is unclear because, apart from the receptor itself, key physiological substrates of the protein tyrosine kinase have not been identified. Insulin alters the level of serine/threonine phosphorylation of several regulatory proteins within minutes, but its ability to promote increased phosphorylation of some proteins and decreased phosphorylation of others is paradoxical², raising the question of whether insulin operates through more than one signalling system.

The stimulation of glycogen synthesis in skeletal muscle by insulin results from the dephosphorylation and activation of glycogen synthase^{3–7}, the rate-limiting enzyme in this pathway. Intravenous administration of insulin reduces the phosphate content of glycogen synthase *in vivo* by ~0.5 mol per subunit within minutes, causing a twofold elevation of activity as tested in the absence (but not in the presence) of the allosteric activator glucose 6-phosphate⁷. Glycogen synthase is phosphorylated on nine serine residues *in vivo* by five or more protein kinases⁸,

THE precise molecular mechanism by which insulin regulates cellular metabolism is unknown. The first step is the interaction of insulin with a receptor bound to the plasma membrane, which

TABLE 1 Phosphorylation of site 1 activates PP1G

Treatment	Phosphate incorporated (mol mol ⁻¹)	Site 1	Site 2	PhP (activity %)	GSP
None	0	0	0	100	100
PK-A phosphorylation	1.8	1.0	0.8	107	236
PP2A dephosphorylation	1.05	0.8	0.25	101	233
AP dephosphorylation	0.10	0.05	0.05	104	115

PP1G (0.7 μ M) in 25 mM Tris-HCl, 25 mM Bistris-HCl pH 7.4 (22 °C) containing 0.1 M NaCl, 0.2 mM EGTA and 1 mM benzamidine was phosphorylated for 20 min at 30 °C with PK-A (40 mU ml⁻¹), 5 mM Mg²⁺ and 0.1 mM ATP in an incubation volume of 0.04 ml. The reaction was stopped by addition of 4 μ l of 0.1 mM PKI and 4 μ l of PP2A (120 mU phosphorylase phosphatase activity ml⁻¹) was added. After a further 20 min, 2 μ l of 0.15 mg ml⁻¹ calf intestinal alkaline phosphatase (AP) from Sigma was added and the incubation continued for a further 60 min. Aliquots withdrawn before and after each treatment were diluted 40-fold into assay buffer⁴³ containing 3 mM okadaic acid to inhibit PP2A (ref. 44), and assayed at a further threefold dilution for phosphorylase phosphatase (PhP) and glycogen synthase phosphatase (GSP)⁴⁵. The presence of 1 nM okadaic acid in the assays has no effect on PP1 (ref. 44). Activity is expressed as a percentage of that measured before phosphorylation. The purified PP1G had a specific activity of 2260 mU mg⁻¹ towards phosphorylase and 27 mU mg⁻¹ towards glycogen synthase (before phosphorylation). No PhP or GSP activity was lost during the experiment in control incubations where PK-A was omitted. Parallel incubations were carried out using [γ -³²P]ATP instead of unlabelled ATP and aliquots were removed and examined for incorporation of ³²P radioactivity into protein⁴⁵ and for phosphorylation of site 1 and site 2 by tryptic phosphopeptide mapping¹³. Phosphorylation of site 1 and site 2 was negligible in the absence of PK-A. Similar results were obtained in two further experiments. ³²P-labelled glycogen phosphorylase phosphorylated to 1.0 mol phosphate per mol subunit with phosphorylase kinase⁴³, ³²P-labelled phosphorylase kinase phosphorylated to 1.6 mol phosphate per mol $\alpha\beta\gamma\delta$ unit with PK-A (ref. 46) and ³²P-labelled glycogen synthase phosphorylated to 1.5 mol phosphate per mol subunit with GSK3 (ref. 46) were prepared as described. The specific radioactivities of each substrate were $\sim 10^6$ c.p.m. nmol⁻¹. PP1G was assayed as described⁴³ at 10 μ M phosphorylase, 1 μ M phosphorylase kinase and 1 μ M glycogen synthase. One unit of protein phosphatase activity is that amount which catalyses the dephosphorylation of 1 μ mol of substrate in 1 min. PK-A was purified as described previously⁴⁷ and one unit of activity was that amount which catalysed the incorporation of 1 μ mol of phosphate into mixed histones in 1 min⁴⁸. Further methodological details are given in Fig. 1.

but the phosphate released in response to insulin is mainly removed from one tryptic peptide containing three phosphoserine residues, termed C30, C34 and C38 (ref. 7). These observations indicate that insulin either inhibits glycogen synthase kinase-3 (GSK3), the protein kinase that phosphorylates C30, C34 and C38 (ref. 9) or activates protein phosphatase-1 (PP1) the principal enzyme that dephosphorylates both glycogen synthase¹⁰⁻¹⁴ and the glycogenolytic enzymes phosphorylase and phosphorylase kinase^{11,15}.

The form of PP1 involved in the regulation of glycogen metabolism (PP1G) is associated with glycogen particles *in vivo* and is far more active in dephosphorylating glycogen synthase (and glycogen phosphorylase) than other forms of PP1, when assayed under physiological conditions¹⁴. It is composed of the catalytic (C) subunit (relative molecular mass 37,000; M_r 37K) complexed to a 160K G subunit responsible for interaction of PP1G with glycogen^{12,13}. The G subunit is phosphorylated by cyclic AMP-dependent protein kinase (PK-A) *in vitro*^{12,16} and in response to adrenalin *in vivo*^{17,18}. Phosphorylation occurs at two serines, site 1 and site 2, that are separated by only 18 residues, and results in dissociation of the C subunit from the G subunit¹⁶. The released C subunit is 5-8-fold less effective than PP1G in dephosphorylating glycogen synthase and glycogen phosphorylase when assayed under physiological conditions¹⁴, and inhibition of PP1 by this means is one of the devices that allows adrenalin to stimulate glycogenolysis and inhibit glycogen synthesis (reviewed in ref. 19).

Dissociation of the G and C subunits correlates with the phosphorylation of site 2, but not site 1, whereas reassociation is promoted by PP2A under conditions that dephosphorylate site 2, but not site 1¹⁶. These findings raised the question of the role of site 1, which is dephosphorylated much more slowly than site 2 by all four main types of protein serine/threonine phosphatase present in eukaryotic cells¹⁶. We now report that site 1 phosphorylation stimulates the dephosphorylation of glycogen synthase and phosphorylase kinase by PP1G. Furthermore, an insulin-stimulated protein kinase present in skeletal muscle phosphorylates site 1 specifically, and the level of phosphorylation of site 1 *in vivo* is increased by insulin. These findings indicate that site 1 phosphorylation underlies the inhibition of glycogenolysis^{20,21} as well as the activation of glycogen synthesis by insulin.

Phosphorylation of site 1 activates PP1G

Although release of the C subunit from PP1G reduces the rate

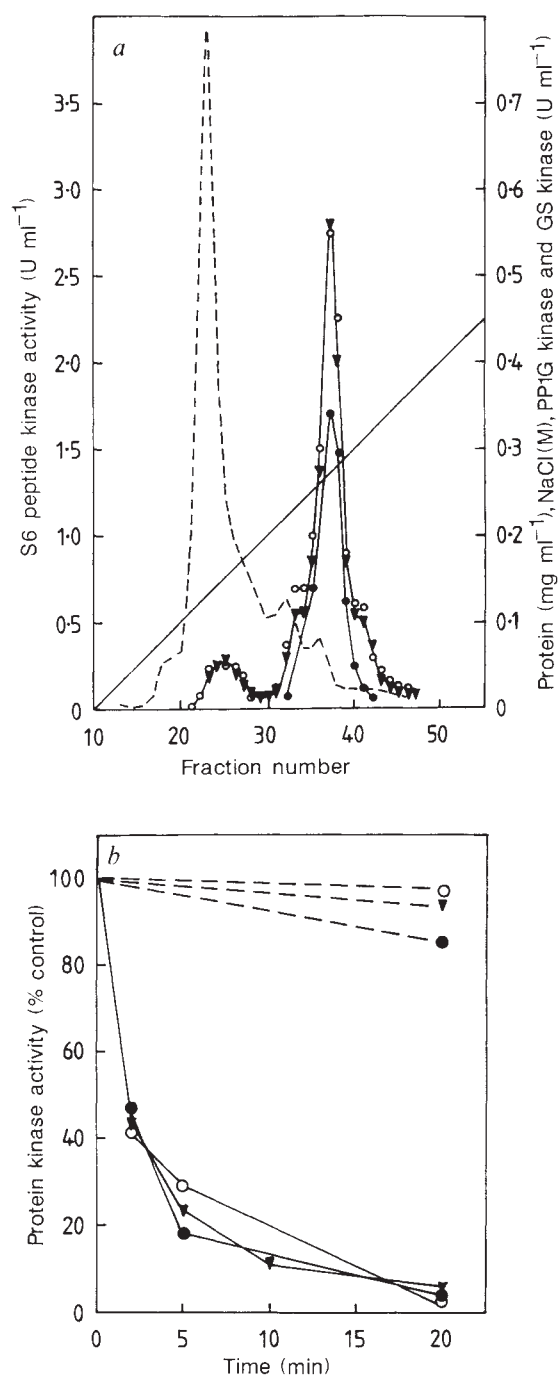
of dephosphorylation of glycogen synthase about fivefold when assayed under physiological conditions (that is, at near physiological ionic strength and under conditions where both PP1G and its substrates are bound to glycogen¹⁴), a 2.5-3-fold increase in activity is observed in the standard assay¹⁴, which is carried out at low ionic strength and in the absence of glycogen. These observations indicate that, in the absence of glycogen, interaction of the G subunit with the C subunit partially suppresses activity towards glycogen synthase, this inhibition being relieved when the C subunit is released from the G subunit. But to our surprise, this enhanced activity towards glycogen synthase was retained following the dephosphorylation of site 2 by PP2A (Table 1) which results in recombination of the C and G subunits¹⁶. Only on subsequent dephosphorylation of site 1 with calf intestinal alkaline phosphatase did the glycogen synthase phosphatase activity return to the level of dephospho-PP1G (Table 1). These experiments indicated that phosphorylation of site 1 is sufficient to stimulate the dephosphorylation of glycogen synthase by PP1G, a finding confirmed by experiments described below using an insulin-stimulated protein kinase that phosphorylates site 1 specifically. In contrast, the dephosphorylation of glycogen phosphorylase by PP1G at low ionic strength and in the absence of glycogen is unaffected by the phosphorylation of site 1 (Table 1).

An insulin-stimulated site-1 kinase

We have identified a protein kinase in skeletal muscle that phosphorylates the G subunit of PP1G at site 1, but not site 2, and partially purified it by successive chromatographic separations on phosphocellulose, DEAE-cellulose and Mono S as described in the legend to Fig. 1. The protein kinase is not specific for the G subunit, and phosphorylates glycogen synthase and several synthetic peptides *in vitro*, the most effective being a 32-residue peptide corresponding to the C-terminus of ribosomal protein S6 (ref. 22). One main peak of protein kinase activity is observed on chromatography on Mono S with either PP1G, glycogen synthase or the S6 peptide (Fig. 1a) at which step it is purified ~ 200 -fold from the phosphocellulose eluate. Several lines of evidence demonstrate that phosphorylation of the G subunit, glycogen synthase and S6 peptide is catalysed by the same protein kinase. First, activity towards each substrate not only copurifies on Mono S, but also during subsequent chromatographies on Mono Q and phenyl-Superose (data not shown). Second, activity towards each substrate is lost in parallel on incubation with PP2A (Fig. 1b), an effect which is abolished

FIG. 1 Chromatography of the ISPK on Mono S and its inactivation by PP2A. *a*, The ISPK purified by chromatography on phosphocellulose and DEAE-cellulose was applied to a 5×0.5-cm column of Mono S equilibrated in buffer C. The column was developed with a 40 ml linear salt gradient from 0–0.5 M NaCl in buffer C, at a flow rate of 1 ml min⁻¹. Fractions of 0.5 ml were collected and assayed for protein kinase activity with 30 μM S6 peptide (▼), 0.25 mg ml⁻¹ PP1G (●) and 0.4 mg ml⁻¹ glycogen synthase (○). Salt gradient, full line and protein⁴⁹, broken line. The recovery of activity varied from 40–70% in different experiments. *b*, Mono S-purified ISPK (1.5 U ml⁻¹ S6 peptide kinase activity) was incubated at 30 °C with the purified catalytic subunit of PP2A (10 mU ml⁻¹, see Table 1 for definition of units) from rabbit skeletal muscle⁴³ in 40 mM MOPS buffer pH 7.0, 1 mM EDTA, 0.1% (v/v) 2-mercaptoethanol. At various times, the reactions were made 0.1 μM in okadaic acid to inactivate PP2A, and 5-μl aliquots were then assayed for ISPK activity using 30 μM S6 peptide (▼), 0.4 mg ml⁻¹ (3.2 μM) glycogen synthase (○) and 0.2 mg ml⁻¹ (1.5 μM) PP1G (●) as substrates. Broken lines, experiments in which the reactions were made 0.1 μM in okadaic acid before the addition of PP2A.

METHODS. Rabbits were injected intravenously with insulin (16.5 μg kg⁻¹) 15 min and 7.5 min before a lethal dose of sodium pentobarbitone. Skeletal muscle from the hind limbs and back was rapidly removed and placed in ice. Subsequent operations were performed at 0–4 °C and all buffers contained 0.1% (v/v) 2-mercaptoethanol, 0.1 mM phenylmethylsulphonyl fluoride, and 1 mM benzimidazole. The muscle was minced, homogenized in the presence of 4 mM EDTA, 5 mM EGTA and 50 mM NaF, and centrifuged (30 min; 4,000g) to pellet the myofibrils. The extract was decanted and 500 ml applied to a 25×5-cm column of phosphocellulose P11 (Whatman) equilibrated in 20 mM MOPS pH 7.0, 4 mM EDTA, 2 mM EGTA, 5% glycerol, 25 mM NaF (buffer A). After washing with buffer A until the A₂₈₀ was below 0.05, the ISPK was eluted (together with GSK3) with 400 ml of buffer A + 0.2 M NaCl. The eluate was concentrated to 100 ml by ultrafiltration, dialysed against 10 mM Tris-HCl pH 7.0 (20 °C), 1 mM EDTA, 10% glycerol, 0.1% Triton X-100 (buffer B) and applied to a 10×5-cm column of DEAE-cellulose (Whatman, DE 52) equilibrated in the same buffer. After washing with buffer B + 0.05 M NaCl to remove all the GSK3, the ISPK was eluted with buffer B + 0.5 M NaCl with a recovery of 80–100%. The active fractions from DEAE-cellulose were concentrated to 40 ml by ultrafiltration and dialysed against 20 mM MOPS pH 6.5, 1 mM EDTA, 0.1% (v/v) 2-mercaptoethanol, 0.1% Triton X-100, 1 mM benzimidazole and 10% (v/v) glycerol (buffer C). ISPK activity was measured at 30 °C in 50-μl incubations containing 8 mM MOPS, 20 mM sodium glycerophosphate, 0.2 mM EDTA, 0.5 mM EGTA, 10 μM PKI (a peptide inhibitor of PK-A (ref. 24), protein substrate and 4 mM magnesium acetate, 0.1 mM [γ -³²P]ATP (0.5 × 10⁶ c.p.m. nmol⁻¹). Okadaic acid (1 μM) was also included when PP1G was the substrate, to prevent autodephosphorylation. Reactions were initiated with Mg-ATP and, for glycogen synthase and PP1G, stopped by addition of 1 ml of 5% (w/v) trichloroacetic acid. Incorporation of phosphate into each substrate was then analysed as described previously⁴⁵. For S6 peptide, reactions were terminated by pipetting 40 μl of the reactions onto small squares of phosphocellulose P81 paper, followed by immersion in 75 mM phosphoric acid (10 ml per square). After washing four times in phosphoric acid and once in acetone, the papers were dried and counted. Phosphorylation was limited to <0.15 mol per mol protein, to ensure that initial rate conditions were met. One unit of ISPK activity was that amount which catalyzed the incorporation of 1.0 nmol of phosphate into each substrate in 1 min. Control incubations carried out in the absence of added substrate showed a negligible incorporation of phosphate into protein. PP1G (ref. 13) and glycogen synthase⁵⁰ were purified from rabbit skeletal muscle and the latter was heated for 10 min at 55 °C in the presence of 5 mM UDP-glucose and 5 mM glucose-6P to inactivate contaminating glycogen synthase kinase activity. S6 peptide was the C-terminal 32-residues of rat liver ribosomal protein S6



(KEAKEKRQEIQAKRRRLSSLRSTSKSGGSQK; single letter amino-acid code). This peptide and PKI were synthesized on an Applied Biosystems peptide synthesizer.

by okadaic acid (Fig. 1b) a potent inhibitor of PP2A (reviewed in ref. 23). Third, the presence of S6 peptide (6 μM) inhibits the phosphorylation of glycogen synthase (2.6 μM) and PP1G (1.8 μM) by 45% and 60%, respectively.

The protein kinase is activated within 15 min of an acute injection of insulin (Table 2), the time at which maximal activation of glycogen synthase is also observed⁷, and for this reason is hereafter referred to as the insulin-stimulated protein kinase (ISPK). Activation by insulin is measured after chromatography on phosphocellulose using the S6 peptide as substrate. The ISPK is the main S6 peptide kinase at this stage (if assayed in the presence of EGTA to inhibit Ca²⁺-dependent protein kinases and the specific peptide inhibitor (PKI) of PK-A, ref. 24),

because no other S6 peptide kinases are removed during the subsequent chromatography on DEAE-cellulose and Mono S. The ISPK is recovered in the 0.05–0.5 M NaCl eluate from DEAE-cellulose with a yield of 80–100%, and only one major peak of S6 peptide kinase activity is observed on chromatography on Mono S, which coelutes with the ISPK (Fig. 1a). The presence of other protein kinases in the phosphocellulose eluate that phosphorylate PP1G and glycogen synthase (for example, GSK3; ref. 25) and the lower activity of the ISPK towards these substrates (Fig. 1a) precludes their use for measuring activation by insulin, nor can activation be determined reliably after chromatography on Mono S because of the variable recovery of activity (40–70%) at this stage.

TABLE 2 Activation of the insulin-stimulated protein kinase by insulin

Preparation	S6 peptide kinase activity (U mg ⁻¹) minus insulin	plus insulin
1	0.026	0.055
2	0.038	0.079
3	0.024	0.086
Average (±s.d.)	0.029 ± 0.006	0.073 ± 0.016

Frozen hind limbs from two animals, one injected with insulin plus propranolol and the other with propranolol alone, were warmed to -10°C and the muscle removed. The ISPK was then partially purified by chromatography on phosphocellulose as described in Fig. 1 with the following modifications: 50 ml of each extract was diluted to 200 ml with buffer A, chromatographed on $8 \times 2\text{-cm}$ columns of phosphocellulose, washed with 200 ml of buffer A + 0.05 M NaCl and eluted with 120 ml of buffer A + 0.2 M NaCl. ISPK activity was determined using the S6 peptide ($30\text{ }\mu\text{M}$) as substrate as described in Fig. 1. Protein was measured by the procedure of Bradford⁴⁹. Stimulation of S6 peptide kinase by insulin was slightly greater than the values shown in the table when activity in the phosphocellulose eluate was normalized to either the protein concentration of the muscle extracts (3.1-fold stimulation) or the glycogen synthase (+glucose-6P) activity (3.5-fold stimulation) of the extracts. Other methods are given in Table 3.

Incubation of partially purified ISPK with PP1G and $\text{Mg}[\gamma\text{-}^{32}\text{P}]\text{-ATP}$ results in phosphorylation of PP1G to a level of $\sim 0.8\text{ mol per mol subunit}$ (Fig. 2a), and digestion with trypsin followed by chromatography on a C_{18} column reveals one major peak of ^{32}P radioactivity coeluting with the tryptic peptide containing site-1 (Fig. 2c). Only traces of phosphate are present in the tryptic peptide containing site 2 (Fig. 2c). The major

^{32}P -labelled peptide also comigrates in isoelectric focussing experiments with the site 1 tryptic phosphopeptide obtained after phosphorylation with PK-A, and when analysed on a gas phase sequencer as described previously²⁶, the ^{32}P radioactivity is entirely released from the peptide after the third cycle of Edman degradation, as expected from previous work²⁷. These observations establish that the ISPK phosphorylates the site 1 phosphoserine, and not any of the other four serine residues in this peptide^{16,27}. Consistent with specific phosphorylation of site 1, the G and C subunits remain fully associated, as $>80\%$ of the activity was recovered in the glycogen pellet in several separate experiments after addition of glycogen (5 mg ml^{-1}) and centrifugation at $150,000\text{g}$ (data not shown).

Phosphorylation of site 1 by the ISPK to 0.8 mol mol^{-1} enhances the glycogen synthase phosphatase activity of PP1G by 2.8-fold in the standard assay (Fig. 2b), confirming the results presented in Table 1 by an independent method. Phosphorylation of site 1 increases the phosphorylase kinase phosphatase activity to a similar extent, but has no effect on the phosphorylase phosphatase activity (Fig. 2b). The near stoichiometric phosphorylation of site 1 by either the ISPK (Fig. 2a) or cAMP-dependent protein kinase¹⁶ and failure of calf intestinal alkaline phosphatase to decrease the glycogen synthase phosphatase activity of purified PP1G (data not shown) indicates that this serine is almost entirely dephosphorylated before phosphorylation.

As the effect of site 2 phosphorylation on PP1G activity is only observed at physiological ionic strength, in the presence of glycogen and under conditions where both PP1G and its substrates are bound to glycogen¹⁴, we examined the effect of

FIG. 2 Activation of PP1G by the insulin-stimulated protein kinase-catalysed phosphorylation of site 1. a, PP1G (0.3 mg ml^{-1} , $\sim 2\text{ }\mu\text{M}$) was phosphorylated at pH 7.4 in the presence (●) and absence (○) of Mono S-purified ISPK (10 mU ml^{-1} of S6 peptide kinase activity, see Fig. 1) using $\text{Mg}[\gamma\text{-}^{32}\text{P}]\text{-ATP}$ as described for phosphorylation by PK-A (ref. 13), except for the addition of $0.1\text{ }\mu\text{M}$ okadaic acid to inhibit autodephosphorylation. b, The effect of phosphorylation on PP1G activity was examined in parallel incubations to a using unlabelled ATP. Aliquots were removed at various times and assayed for phosphorylase phosphatase (PhP, ▼), phosphorylase kinase phosphatase (PhKP, ○) and glycogen synthase phosphatase (GSP, ●) activities at a 300-fold final dilution to relieve inhibition by okadaic acid⁴⁴. The specific activities of PP1G towards phosphorylase, phosphorylase kinase and glycogen synthase (before phosphorylation) were $2,160\text{ mU mg}^{-1}$, 60 mU mg^{-1} and 21 mU mg^{-1} , respectively. c, PP1G from a containing 0.8 mol per mol phosphate was precipitated with trichloroacetic acid, digested with trypsin and chromatographed on a Vydac C_{18} column in the presence of 0.1% trifluoroacetic acid, using the protocol that prevents any loss of the hydrophobic site-2 tryptic phosphopeptide¹³. Solid line, ^{32}P radioactivity in arbitrary units (detected with an on-line monitor). Broken line, the acetonitrile gradient. The positions of the tryptic peptides containing site 1 and site 2 are marked. Other methods are given in Table 1.

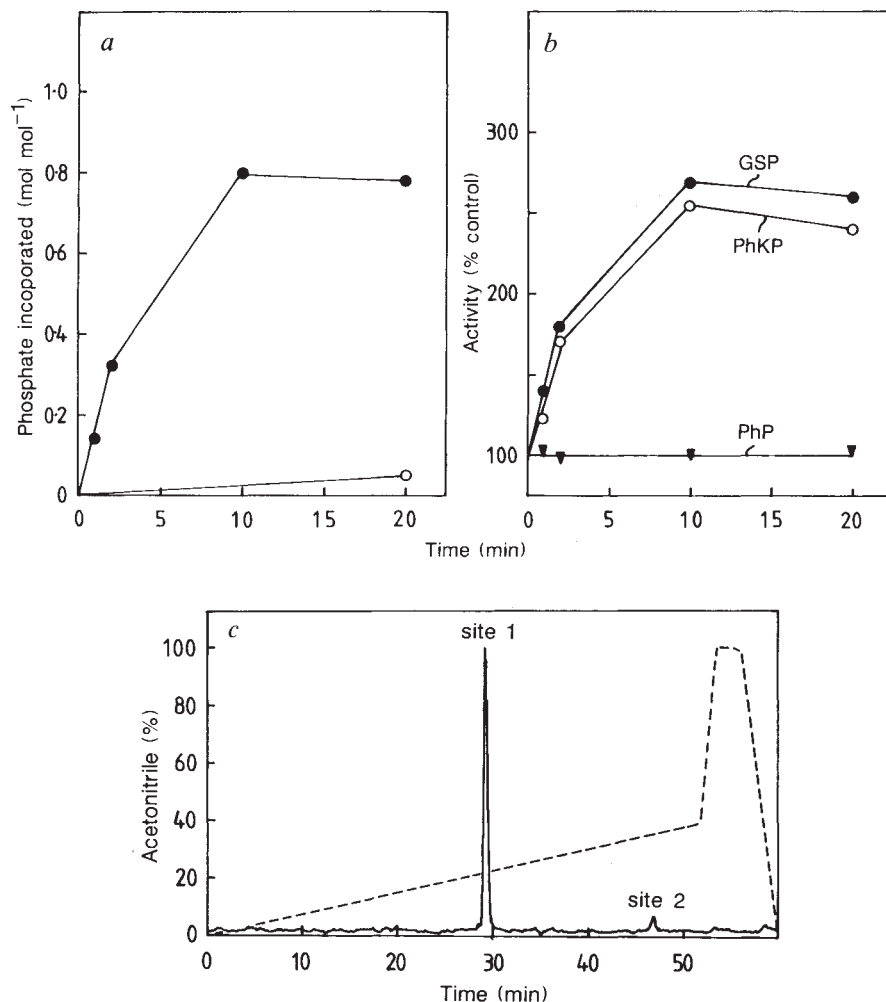


TABLE 3 *In vivo* phosphorylation of the G subunit

Preparation	Injection	Site-1	Site-2
1	propranolol	0.26 ± 0.01	0.16 ± 0.02
2	propranolol	0.30 ± 0.03	0.21 ± 0.02
Average of 1 and 2	propranolol	0.28 ± 0.02	0.19 ± 0.03
3	propranolol + insulin	0.53 ± 0.02	0.20 ± 0.05
4	propranolol + insulin	0.48 ± 0.01	0.19 ± 0.02
Average of 3 and 4	propranolol + insulin	0.51 ± 0.02	0.20 ± 0.01

PP1G isolated in the presence of phosphatase inhibitors from the muscle homogenates of rabbits injected with the β -adrenergic antagonist propranolol, or propranolol plus insulin^{17,18}, was digested with trypsin and the tryptic peptides containing site 1 and site 2, purified by gel filtration on Sephadex G50 Superfine and chromatography on a Vydac C18 reverse phase column under conditions that do not resolve the phospho and dephosphopeptides^{17,18}. The phospho and dephospho forms were then separated by reverse phase chromatography at pH 6.5, and identified by fast atom bombardment mass spectrometry and amino-acid sequencing. Phosphorylation stoichiometries were determined by quantitative amino acid analyses of the separated phospho- and dephosphopeptides as well as by integration of the areas under the ultraviolet absorbance peaks corresponding to the phospho- and dephosphopeptides^{17,18}. The results obtained by each procedure were then averaged to yield the values shown. Rabbits (about 4 kg) that had been deprived of food for 24 h were anaesthetized by intravenous injection of Sagatal (50 mg kg⁻¹) (May and Baker, Dagenham, UK). The skin of both hind limbs was reflected and a ligature placed around each femoral artery. Propranolol (2 mg kg⁻¹) and insulin (33 μ g kg⁻¹) were administered through the ear vein 15 min before freeze-clamping and a further injection of insulin was made 7.5 min before freeze clamping. Control animals received propranolol but no insulin. The hind limbs were freeze-clamped using steel plates precooled to the temperature of liquid nitrogen and the animals were then injected with a lethal dose of sodium pentobarbitone. The limbs were removed at the level of the head of the femur after tying off the blood vessels, and kept at -70 °C. The frozen hind limbs from three animals were warmed to -10 °C and the muscle excised.

site 1 phosphorylation in this 'physiological assay'. These experiments were carried out as described¹⁴, except that glycogen synthase and phosphorylase kinase were incubated with glycogen for 24 h at 0 °C before the assays and were 100% bound and 50% bound to glycogen, respectively. Six experiments on two PP1G preparations showed that phosphorylation of site 1 to 0.58 ± 0.02 mol mol⁻¹, increased the glycogen synthase phosphatase activity of PP1G by 2.24 ± 0.17 times in the 'standard assay' and 1.91 ± 0.20 times in the physiological assay (s.d., $n = 6$). The phosphorylase kinase phosphatase activity was increased 2.15 ± 0.17 times in the standard assay and 1.86 ± 0.15 times in the physiological assay (s.d., $n = 6$). Thus, in contrast to site 2, the effects of site 1 phosphorylation on activity are similar in the standard and physiological assays.

The native 160K G subunit is extremely sensitive to proteolysis and is partially degraded during isolation of PP1G, to fragments of 103K and smaller^{13,27}. Although degradation of the G subunit to fragments as small as 40K does not affect interaction with the C subunit or with glycogen, nor remove the phosphorylation sites^{13,28}, it was important to demonstrate that PP1G containing intact 160K subunit, is also activated by the ISPK. Freshly isolated glycogen particles were therefore digested with α -amylase, centrifuged for 1 h at 100,000g to pellet insoluble material and the supernatant removed and incubated with the ISPK and Mg-ATP as in Fig. 2. Stimulation of the glycogen synthase phosphatase (2.7 ± 0.1 times) and phosphorylase kinase phosphatase (2.4 ± 0.1 times) activities (average of duplicate determinations on two preparations) was similar to that obtained with purified PP1G. Immunoblotting experiments carried out before and after phosphorylation established that the G subunit was present as the undegraded 160K species throughout the experiment (data not shown), as expected from previous results¹³.

Site 1 phosphorylation *in vivo*

We have developed procedures for measuring the *in vivo* phosphorylation state of site 1 and site 2 on the G subunit^{17,18}. Using the same methodology, we now find that an acute injection of insulin into rabbits starved for 24 h increases the phosphorylation of site 1 to 0.51 mol mol⁻¹, as compared to 0.28 mol mol⁻¹

in control preparations where insulin is omitted (Table 3). In contrast, the level of phosphorylation of site 2 (0.19 mol mol⁻¹ in control animals, 0.20 mol mol⁻¹ after injection of insulin) is unaffected by insulin (Table 3). The level of phosphorylation of site 2 in the absence or presence of insulin is similar to the values we reported previously for fed animals¹⁸, but site 1 phosphorylation in the absence of insulin is lower^{17,18}. This difference is explained by a methodological improvement introduced in the present study in which the hind limbs of anaesthetized animals were freeze-clamped 15 min after injection of insulin to prevent post mortem increases in the phosphorylation state of site 1 during excision of the muscle (Table 3). If the muscle is removed without freeze-clamping, the level of site 1 phosphorylation is ~ 0.6 mol mol⁻¹ (phospho site 1 : dephospho site 1 = 1.7) in control animals increasing to ~ 0.85 mol mol⁻¹ (phospho site 1 : dephospho site 1 = 5.1) in animals injected with insulin (data not shown).

Discussion

Here, we have demonstrated that phosphorylation of the G subunit at site 1 increases the rate of dephosphorylation (activation) of glycogen synthase and the rate of dephosphorylation (inhibition) of phosphorylase kinase by PP1G (Fig. 2b, Table 1). Also, that an insulin-stimulated protein kinase present in skeletal muscle phosphorylates site 1 specifically (Fig. 2c), and that site 1, but not site 2, is phosphorylated *in vivo* in response to insulin (Table 3). These observations suggest that phosphorylation of site 1 by the ISPK underlies the activation of glycogen synthesis and inhibition of glycogenolysis^{20,21} by insulin (Fig. 3).

The ISPK phosphorylates glycogen synthase and the G-subunit at similar rates (Fig. 1), and we have identified the phosphorylation site as N7. Following tryptic digestion of ³²P-labelled glycogen synthase, all the radioactivity is recovered in a trichloroacetic acid-insoluble peptide, which after dissolution in 0.1 M NH₄HCO₃ elutes from Sephadex G50 as a single peak

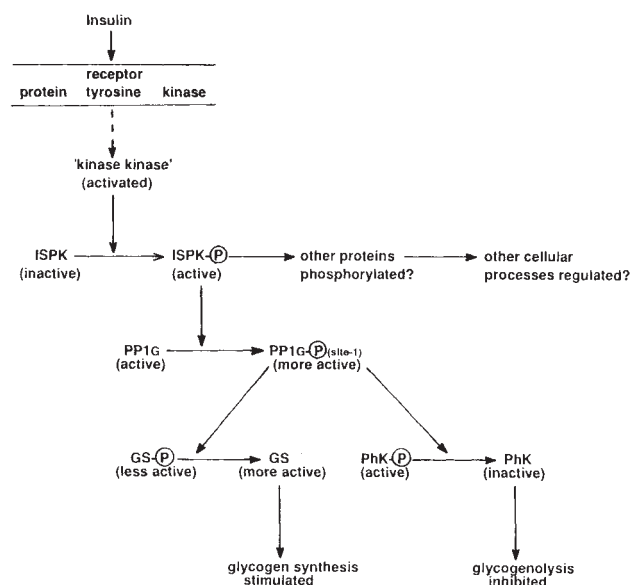


FIG. 3 Molecular mechanism by which insulin regulates glycogen metabolism in mammalian skeletal muscle. The interaction of insulin with its receptor activates its protein tyrosine kinase activity, leading to activation of a protein Ser/Thr 'kinase kinase' by an unknown mechanism. The latter phosphorylates and activates the insulin-stimulated protein kinase (ISPK), allowing it to phosphorylate site 1 on the G subunit of PP1G. Phosphorylation of site 1 stimulates the dephosphorylation of residues C30, C34 and C38 on glycogen synthase (GS) and the β -subunit of PhK, activating and inactivating these two enzymes, respectively. As a result, glycogen synthesis is stimulated and glycogenolysis inhibited. The ISPK phosphorylates other proteins *in vitro* and therefore has the potential to mediate other actions of insulin *in vivo*.

comigrating with the tryptic peptide N5-N38 (ref. 29). The identity of the peptide has been confirmed by sequence analysis and the ^{32}P radioactivity is entirely released after the third cycle of Edman degradation corresponding to N7. If the ISPK phosphorylates N7 *in vivo*, this may explain why we did not observe significant dephosphorylation of this serine residue *in vivo* in response to insulin⁷, despite the fact that N7 is dephosphorylated as efficiently as C30, C34 and C38 by PP1G *in vitro*⁷. The other three serines that are phosphorylated *in vivo* in the absence of adrenergic stimulation (C42, C46 and C100)⁸ are all rather resistant to dephosphorylation *in vitro*^{8,29}, explaining why the effect of insulin *in vivo* is largely confined to C30, C34 and C38.

An intriguing aspect of this control system is that adrenalin (acting through PK-A) and insulin (acting through the ISPK), which have antagonistic effects on glycogen metabolism, both promote the phosphorylation of site 1. However, the activating effect of site 1 phosphorylation would be overridden during adrenergic stimulation by the phosphorylation of site 2, which dissociates the C subunit from the G subunit. After adrenergic stimulation ceases, the rapid dephosphorylation of site 2 would lead to recombination of the G and C subunits, allowing glycogen to be resynthesized more rapidly as a result of the phosphorylation of site 1. Thus, adrenalin not only inactivates PP1G through the phosphorylation of site 2, but primes PP1G to resynthesize glycogen very rapidly after adrenergic stimulation has terminated by phosphorylating site 1.

The opposing effects of site 1 and site 2 phosphorylation on PP1G activity represents an unusual situation, where one phosphorylation activates and another inactivates the same enzyme. To our knowledge, there are only three other examples of this phenomenon. The protein tyrosine kinase activity of the EGF receptor is inhibited by the protein kinase C-catalysed phosphorylation of threonine 654 (ref. 30), whereas autophosphorylation of the receptor on tyrosine residues causes a slight enhancement of activity towards some substrates³¹. The cellular homologue of the transforming protein of Rous sarcoma virus, p60^{c-src}, another protein tyrosine kinase, is inhibited by phosphorylation of tyrosine 527, but seems to be activated by autophosphorylation of tyrosine 416 (ref. 32). Acetyl CoA carboxylase, the rate-limiting enzyme in fatty-acid synthesis, is inactivated by the AMP-activated protein kinase that phosphorylates serine 79 (ref. 33), whereas phosphorylation of a different serine residue by an insulin-stimulated protein kinase from adipose tissue relieves the effects of an unidentified inhibitory substance³⁴.

Several protein serine/threonine kinases have been identified that are stimulated by insulin and other growth factors³⁵, raising the question of their relationship to the ISPK. Two insulin-stimulated protein kinases, the 70K and 90K S6 kinases, that resemble the ISPK in their phosphorylation of synthetic peptides corresponding to the C-terminus of ribosomal protein S6 and their inactivation by PP2A have been described in other cells and tissues [refs 22, 36, 37]. We have therefore compared the specificities of these enzymes and find that the substrate preference of the ISPK is quite different from that of the 70K S6 kinase from rat liver (provided by J. Avruch²²). The latter enzyme is at least tenfold more active towards 40S ribosomal subunits and tenfold less active towards the G subunit of PP1G when activities are matched for activity towards the 32-residue S6 peptide. On the other hand, the 90K S6 kinase from *Xenopus* oocytes (provided by J. Maller, also called S6 kinase-II, ref. 37) and the ISPK phosphorylate the 32-residue S6 peptide, the peptide LRRASLG ('Kemptide'), 40S ribosomal subunits and the G subunit of PP1G at similar relative rates. Site 1 is also the major residue on the G subunit phosphorylated by the 90K S6 kinase-II, but more phosphate is introduced into site 2 than with the ISPK. After phosphorylation of the G subunit to 0.77 mol mol⁻¹ by the 90K S6 kinase 0.63 mol mol⁻¹ was incorporated into site 1 and 0.14 mol mol⁻¹ into site 2. In contrast, after phosphorylation to 0.82 mol mol⁻¹ by the ISPK, 0.75 mol

mol⁻¹ was incorporated into site 1 and 0.07 mol mol⁻¹ into site 2. Another difference is that the 90K S6 kinase-II consistently phosphorylates glycogen synthase two times more rapidly than the ISPK (when activities are normalized to S6 peptide kinase activity), although both protein kinases phosphorylate the same residue on glycogen synthase (N7, ref. 38). These results suggest that the ISPK may be related to the family of 90K S6 kinases, but detailed structural analysis will be required to establish whether this is so.

The finding that the ISPK is inactivated by PP2A (Fig. 2), a protein serine/threonine phosphatase, and that sodium fluoride (a protein serine/threonine phosphatase inhibitor) must be present to preserve activity during purification (unpublished observations), indicates that the ISPK is only active when it is phosphorylated on a serine or threonine residue(s). Insulin is therefore likely to exert its effect by promoting phosphorylation of the ISPK. Following inactivation by PP2A, the ISPK cannot be reactivated by preincubation with Mg-ATP, implying that a distinct protein serine/threonine 'kinase kinase' is required (Fig. 3). The 90K S6 kinase-II can be reactivated up to 30% by another growth factor-stimulated protein kinase, termed MAP2 kinase³⁷, but whether this enzyme is responsible for activation *in vivo* is unknown.

The broad specificity of the ISPK *in vitro* is potentially important, because it raises the possibility that this enzyme may have other physiological substrates and mediate further actions of insulin that result from increased serine/threonine phosphorylation (Fig. 3), such as glucose transport³⁹. Most importantly, however, the activation of a protein phosphatase (PP1G) by the ISPK explains how insulin can stimulate the phosphorylation of some proteins and the dephosphorylation of others using a common mechanism (Fig. 3).

PP1G is just one of many forms of PP1 in mammalian cells in which the C subunit is complexed to different (regulatory) proteins (reviewed in refs 19, 40). The activation of PP1G by insulin does not exclude the possibility that other forms of PP1 are stimulated by this hormone and play a part in the acute regulation of different cellular processes by insulin. Indeed, PP1 activity increases by 40% following a 5-min exposure of 3T3 cells to insulin and growth factors⁴¹, and acute effects of insulin on PP1 activity in the liver have also been reported (for example, ref. 42, reviewed in ref. 19). But in these studies PP1 was assayed by its ability to dephosphorylate glycogen phosphorylase^{41,42}. The mechanism by which insulin activates other forms of PP1 must therefore differ from that described here, because phosphorylation of the G subunit at site 1 has no effect on the rate of dephosphorylation of phosphorylase by PP1G (Fig. 2b, Table 1). □

Received 30 July; accepted 11 October 1990.

1. Zick, Y. *Crit. Rev. Biochem. molec. Biol.* **24**, 217-269 (1989).
2. Denton, R. M. *Adv. Cyclic Nuc. Prot. Phosph. Res.* **20**, 293-401 (1986).
3. Villar-Palasi, C. & Larner, J. *Biochim. biophys. Acta* **39**, 171-173 (1960).
4. Danforth, W. H. *J. Biol. Chem.* **240**, 588-593 (1965).
5. Roach, P. J., Rosell-Perez, M. & Larner, J. *FEBS Lett.* **80**, 95-98 (1977).
6. Uhing, R. J., Shikama, H. & Exton, J. H. *FEBS Lett.* **134**, 185-188 (1981).
7. Parker, P. J., Caudwell, F. B. & Cohen, P. *Eur. J. Biochem.* **130**, 227-234 (1983).
8. Poulter, L. et al. *Eur. J. Biochem.* **175**, 497-510 (1988).
9. Rylatt, D. B., Aitken, A., Billham, T., Condon, G. D. & Cohen, P. *Eur. J. Biochem.* **107**, 529-537 (1980).
10. Ingebritsen, T. S., Stewart, A. A. & Cohen, P. *Eur. J. Biochem.* **132**, 297-307 (1983).
11. Alemany, S., Tung, H. Y. L., Shenolikar, S., Pilakis, S. J. & Cohen, P. *Eur. J. Biochem.* **145**, 51-56 (1984).
12. Strålfors, P., Hiraga, A. & Cohen, P. *Eur. J. Biochem.* **149**, 295-303 (1985).
13. Hubbard, M. J. & Cohen, P. *Eur. J. Biochem.* **180**, 457-465 (1989).
14. Hubbard, M. J. & Cohen, P. *Eur. J. Biochem.* **186**, 711-716 (1989).
15. Ingebritsen, T. S., Foulkes, J. G. & Cohen, P. *Eur. J. Biochem.* **132**, 263-274 (1983).
16. Hubbard, M. J. & Cohen, P. *Eur. J. Biochem.* **186**, 701-709 (1989).
17. MacKintosh, C., Campbell, D. G., Hiraga, A. & Cohen, P. *FEBS Lett.* **234**, 189-194 (1988).
18. Dent, P., Campbell, D. G., Caudwell, F. B. & Cohen, P. *FEBS Lett.* **259**, 281-285 (1990).
19. Cohen, P. *A. Rev. Biochem.* **58**, 453-508 (1989).
20. Zhang, J. N., Hiken, J., Davis, A. E. & Lawrence, J. C. *J. Biol. Chem.* **264**, 17513-17523 (1989).
21. Challiss, R. A. J., Crabtree, B. & Newsholme, E. A. *Eur. J. Biochem.* **163**, 205-210 (1987).
22. Price, D. J., Nemenoff, R. A. & Avruch, J. *J. Biol. Chem.* **264**, 13825-13833 (1989).
23. Cohen, P., Holmes, C. F. B. & Tsukitani, Y. *Trends Biochem. Sci.* **15**, 98-102 (1990).
24. Scott, J. D., Glaccum, M. B., Fischer, E. H. & Krebs, E. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1613-1616 (1986).
25. Dent, P., Campbell, D. G., Hubbard, M. J. & Cohen, P. *FEBS Lett.* **248**, 67-72 (1989).
26. Wang, Y. et al. *Analyt. Biochem.* **174**, 537-547 (1987).

27. Caudwell, F. B., Hiraga, A. & Cohen, P. *FEBS Lett.* **194**, 85–90 (1988).
28. Hiraga, A., Kemp, B. E. & Cohen, P. *Eur. J. Biochem.* **183**, 253–258 (1989).
29. Embi, N., Parker, P. J. J. & Cohen, P. *Eur. J. Biochem.* **115**, 405–413 (1981).
30. Davis, R. J. *J. Biol. Chem.* **263**, 9462–9469 (1988).
31. Downward, J. D., Waterfield, M. D. & Parker, P. J. *J. Biol. Chem.* **260**, 14538–14546 (1985).
32. Reuter, C., Findik, D. & Presek, P. *Eur. J. Biochem.* **190**, 343–350 (1990).
33. Hardie, D. G., Carling, D. & Sim, A. T. R. *Trends Biochem. Sci.* **14**, 20–23 (1989).
34. Borthwick, A. C., Edgell, N. J. & Denton, R. M. *Biochem. J.* **270**, 795–801 (1990).
35. Czech, M. P., Klariand, J. K., Yagaloff, K. A., Bradford, A. P. & Lewis, R. E. *J. Biol. Chem.* **263**, 11017–11020 (1988).
36. Ballou, L. M., Siegmund, M. & Thomas, G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7154–7158 (1988).
37. Sturgill, T. W., Ray, L. B., Erikson, E. & Maller, J. L. *Nature* **334**, 715–718 (1988).
38. Erikson, E. & Maller, J. *Second Mess. Phosphoproteins* **12**, 135–143 (1988).
39. Haystead, T. et al. *Nature* **337**, 78–81 (1989).
40. Cohen, P. & Cohen, P. T. W. *J. Biol. Chem.* **264**, 21435–21438 (1989).
41. Chan, C. P., McNall, S. J., Krebs, E. G. & Fischer, E. H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6257–6261 (1988).
42. Toth, B., Bollen, M. & Stalmans, W. *J. Biol. Chem.* **263**, 14061–14066 (1988).
43. Cohen, P. et al. *Meth. Enzym.* **159**, 390–408 (1988).
44. Cohen, P., Klumpp, S. & Schelling, D. L. *FEBS Lett.* **250**, 596–600 (1989).
45. Guy, P. S., Cohen, P. & Hardie, D. G. *Eur. J. Biochem.* **115**, 399–405 (1981).
46. Stewart, A. A., Hemmings, B. A., Cohen, P., Goris, J. & Merlevede, W. *Eur. J. Biochem.* **115**, 197–205 (1981).
47. Reimann, E. M. & Beham, R. A. *Meth. Enzym.* **99**, 51–55 (1983).
48. Cohen, P., Nimmo, G. A. & Antoniwi, J. F. *Biochem. J.* **162**, 435–444 (1977).
49. Bradford, M. M. *Analyt. Biochem.* **72**, 248–252 (1976).
50. Nimmo, H. G., Proud, C. G. & Cohen, P. *Eur. J. Biochem.* **68**, 21–30 (1976).

ACKNOWLEDGEMENTS. We thank D. Campbell for protein sequencing and peptide synthesis, C. MacKintosh and M. Hubbard for constructive advice, Y. Tsukitani (Fujiwara Pharmaceutical Company, Tokyo) for a gift of okadaic acid and the Medical Research Council, Royal Society and The Fondation pour la Recherche Medicale (A.L.) for financial support. P.D. is the recipient of a postgraduate studentship from the Wellcome Trust. The first three authors made equally important contributions to this work.

Targets of homeotic gene control in *Drosophila*

Alex P. Gould, Jenny J. Brookman, David I. Strutt & Robert A. H. White

Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK

The homeotic gene *Ultrabithorax* (*Ubx*) encodes homeodomain-containing transcription factors that determine segmental identity in *Drosophila*. Here, an immunopurification procedure is described that enriches for embryonic chromatin fragments containing binding sites for *Ubx* protein. In two cases these binding sites are located near embryonic transcription units regulated by the *Ubx* locus *in vivo*. Thus, these transcripts may correspond to *Ubx* target genes involved in elaborating segment-specific developmental pathways.

HOMEOTIC selector genes specify segmental differences in *Drosophila*. The cells of a particular segment follow a specific developmental pathway because of the combination of homeotic genes that they express^{1,2}. Homeotic genes are thought to control developmental pathways by regulating subordinate target genes responsible for segment-specific morphogenesis³. There is much support for this hypothesis. Homeotic genes encode homeodomain-containing proteins (reviewed in refs 4 and 5) that exhibit sequence-specific DNA binding^{6–8} and can act as transcriptional regulators^{9–13}. Furthermore, homeotic genes regulate the expression of other homeotic genes^{14,15} and some are autoregulated^{16,17}. But despite evidence for the regulatory capabilities of homeotic gene products, there has been little progress in identifying the subordinate target genes. The identification of these target genes is essential for understanding how combinations of homeotic gene products are used to specify individual segmental pathways of development and also what morphogenetic mechanisms are used to generate segmental diversity.

On the basis of the specificity of DNA-binding *in vitro*, it has been estimated that a homeodomain protein could bind to several thousand sites in the *Drosophila* genome¹⁸. As it is unlikely that all these sites have functional significance, a more directed approach than binding *in vitro* is required to identify targets used *in vivo*. Also, interactions with other proteins are probably important in determining target-site specificity in the nucleus^{19,20}. For these reasons, we have developed a method for the isolation of DNA sequences bound by homeotic gene products in chromatin isolated from *Drosophila* embryos²¹. This method has enabled us to identify at least two transcription units that are regulated by homeotic genes *in vivo* and that are good candidates for subordinate target genes under the direct control of the homeotic gene *Ultrabithorax* (*Ubx*). We expect that the method will be of general application to the isolation

of *in vivo* targets of DNA-binding proteins, including those from organisms where a comprehensive genetic analysis is not possible.

The immunopurification strategy

The starting point for the isolation of sites bound *in vivo* by homeotic gene products was the availability of a specific mono-

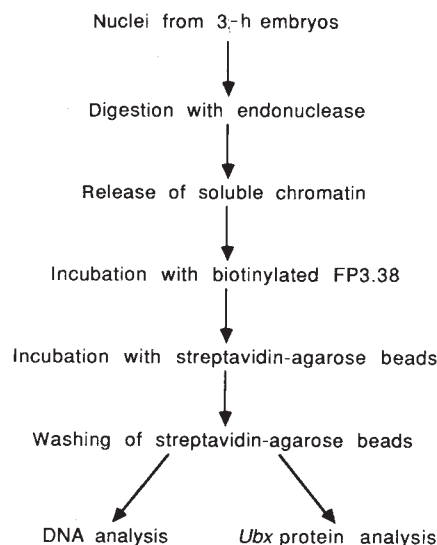


FIG. 1 The immunopurification strategy. FP3.8 is a monoclonal antibody directed against *Ubx* proteins²².

METHODS. Embryonic nuclei⁴⁶ (3×10^7) were resuspended in 10 mM HEPES buffer pH 7.9, 10 mM KCl, 2 mM MgCl₂, 0.5 mM DTT, 15 $\mu\text{g ml}^{-1}$ aprotinin, 5 $\mu\text{g ml}^{-1}$ leupeptin, 2 $\mu\text{g ml}^{-1}$ pepstatin and 0.5 mM PMSF. Nuclei were digested, either by adding 1 mM MgCl₂ and 10⁴ Units of *Hae*III and incubating for 30 min, or by adding 0.5 mM CaCl₂ and 16 Units of micrococcal nuclease (Sigma) followed by 5 min incubation on ice, terminated with 2.5 mM EGTA. Nuclei were centrifuged at 1,500g for 2 min, and the pellet resuspended in TEP (12 mM Tris-base neutralized to pH 7.5 with 3 mM EDTA free acid, 15 $\mu\text{g ml}^{-1}$ aprotinin, 5 $\mu\text{g ml}^{-1}$ leupeptin, 2 $\mu\text{g ml}^{-1}$ pepstatin and 0.5 mM PMSF) to release soluble chromatin. Following 15 min incubation and centrifugation at 1,500g for 5 min, NaCl and BSA were added to the released supernatant to 100 mM and 1 mg ml⁻¹, respectively. After centrifugation at 130,000g for 20 min, the supernatant was preabsorbed for 20 min with 200 μl of streptavidin-agarose beads (Sigma). Preabsorbed chromatin was incubated either with (+antibody) or without (–antibody) 10 $\mu\text{g ml}^{-1}$ biotinylated FP3.8 for 45 min and then with 40 μl of streptavidin-agarose for 30 min. Beads were washed in TEP + 100 mM NaCl before either protein (see Fig. 2) or DNA analysis. Detailed descriptions of the immunopurification method and DNA analysis will be published elsewhere (A.G., J.B., D.S. and R.W., manuscript in preparation).

clonal antibody directed against the products of the *Ubx* gene²². The immunopurification method is outlined in Fig. 1. In brief, a soluble chromatin fraction is prepared from embryonic nuclei by subjecting endonuclease-treated nuclei to a low salt/EDTA lysis procedure²³. From the soluble fraction, which contains at least 50% of the total DNA and *Ubx* protein, about 50% of the remaining *Ubx* protein can be specifically immunopurified (Fig. 2).

DNA in the immunopurified fraction, after *Hae*III nuclear digestion, was cloned into vector M13 for sequencing. About twice the number of clones were obtained from the DNA following (+antibody) immunopurification as from the (–antibody) control purification. In contrast to this small bulk DNA enrichment, the immunopurification results in a marked enrichment for specific DNA sequences. This enrichment was seen on screening for matches to a homeodomain-binding sequence, defined *in vitro*. Many homeodomains will bind with high affinity to DNA containing matches to either the consensus (TAA)_n or TCAATTAAAT^{7,24}. It is not clear whether these two consensus sequences represent different classes of binding sites. The *fushi tarazu* homeodomain (which shares the same residues in the third helix of the homeodomain with *Ubx*, *Antennapedia* and *Deformed*)⁵ shows comparable binding to both consensus classes²⁴. Indeed, both consensus sequences have a central TAAT core that seems critical for *Antennapedia* and *Deformed* homeodomain binding^{5,8}. We screened sequences from the +antibody and –antibody clones for matches to a weight matrix²⁵ constructed from homeodomain-binding sequences of the TCAATTAAAT class²⁴. The +antibody sequences gave 13 matches (within two standard deviations of the weight-matrix mean²⁵) in 9.8 kilobases (kb) of sequence, whereas in the –antibody sequences, no matches were found in 6.5 kb. Thus, immunopurification of chromatin with the anti-*Ubx* monoclonal antibody results in an enrichment for sequences predicted to bind *Ubx* protein *in vitro*.

To assess whether these matches to the weight matrix correlated with *Ubx* protein-binding sites, four +antibody clones

containing high-scoring matches (clones 9, 35, 48 and 72) and three –antibody control clones were assayed for *Ubx* protein binding *in vitro* (Fig. 3). Selective binding, under conditions of high non-specific competitor concentration, was observed for two of the +antibody clones (35 and 48). In contrast, 9 and 72 and the three –antibody clones showed little evidence of significant binding to *Ubx* protein. Thus, a subset of clones containing sequence matches to the weight matrix are able to bind *Ubx* protein with high selectivity *in vitro*.

Transcript distributions

Genomic DNA immediately surrounding the four high-scoring +antibody clones was screened for the presence of transcription units that might be regulated by the *Ubx* locus. Labelled restriction fragments from phage λ genomic clones surrounding the 9, 35, 48 and 72 M13 clones were used in an *in situ* hybridization screen for the presence of embryonic transcripts. In all four cases, transcripts were detected with expression patterns that

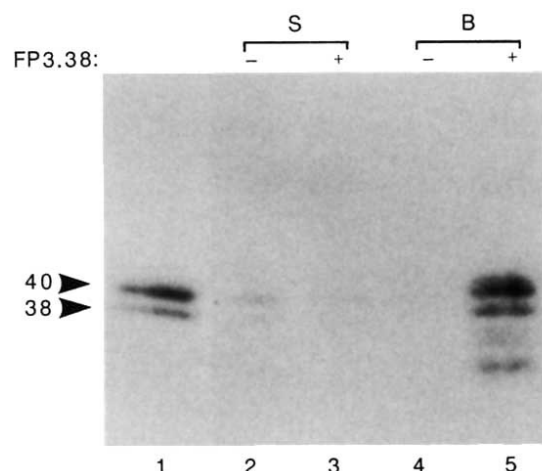


FIG. 2 Immunopurification of *Ubx* proteins from soluble chromatin. Protein blot analysis of the *Ubx* gene products remaining in the supernatant (S, lanes 2 and 3) and bound to the streptavidin-agarose beads (B, lanes 4 and 5) after an immunopurification with FP3.38 (+, lanes 3 and 5), or in a control purification without antibody (–, lanes 2 and 4). Lane 1 shows a total nuclear extract from the initial preparation of nuclei used in this experiment. The sizes of the major *Ubx* protein species (K) are indicated. The brackets indicate minor protein species which may correspond to *Ubx* protein degradation products and are not detected in all immunopurifications. Note that lanes 4 and 5 contain 50% of the total protein extracted from the beads, whereas the supernatant fractions in lanes 2 and 3 contain only 10% of the total reaction.

METHODS. The immunopurification, using micrococcal nuclease digestion, was performed as described in the legend to Fig. 1. *Ubx* proteins were detected by protein blotting as described²².

a	+antibody		Weight-matrix	Cytological
	clone	Size (bp)	matches	location
	48	293	TCAATTAAAT	97D
			TCAATTAGCC	
	9	131	TTAATTAAAT	32D
			CAAATTAAAT	
	35	110	TCAATTAAAC	64C
	72	71	TAAATTAAAT	67C

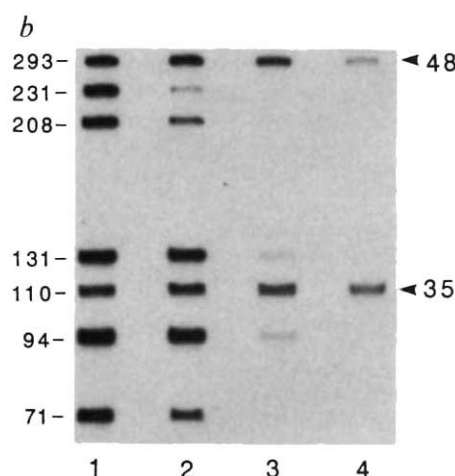


FIG. 3 Selective binding of +antibody clones to *Ubx* protein *in vitro*. a, Sizes, weight-matrix matches and cytological locations⁴⁷ of the four +antibody clones. b, Lane 1 shows the input DNA ladder (size in base pairs (bp) indicated on the left). The three –antibody control inserts have sizes of 231, 208 and 94 bp. Lanes 2–4 show the inserts bound in the presence of 0, $\sim 3 \times 10^3$ and $\sim 10^4$ fold excesses, respectively, of salmon sperm competitor DNA. Only the 35, and 48 inserts show selective binding in the presence of high concentrations of salmon sperm DNA.

METHODS. A heparin-agarose purified *Ubx* fusion protein extract (50 μ g) (R. Weinzierl and R.W., manuscript in preparation) was incubated in 50 μ l of binding buffer (50 mM Tris-HCl pH 7.6, 180 mM KCl, 0.1% Triton X-100, 5 mM β -mercaptoethanol, 10% glycerol, 50 μ g ml^{–1} BSA, 0.6 mg ml^{–1} poly(dI).poly(dC)) for 20 min at 4 °C. ³²P-labelled insert DNA (0.3 ng per clone) and 0, 0.8, or 4 μ g of sonicated salmon sperm DNA were added and incubation continued for 45 min. The binding reaction was then added to 800 μ g of anti-mouse IgG conjugated Dynabeads-M280 (Dyna) saturated with FP5.7, a monoclonal antibody directed against *Ubx* proteins (R. Weinzierl and R.W., manuscript in preparation). After 90 min incubation, beads were washed twice in binding buffer and the extracted DNA was electrophoresed in a denaturing polyacrylamide gel⁴⁸.

are suggestive of homeotic gene regulation; they show segment-to-segment modulation, which is only apparent after the onset of homeotic protein expression.

In two cases (35 and 48) the transcript patterns support a regulation by *Ubx* (Fig. 4). Transcripts detected by probes in the 35 region (called, 35 transcripts) show a segment-to-segment modulated pattern in the central nervous system (CNS) of shortening germ-band embryos (stage 12, staging according to ref. 26) and later embryos (Fig. 4a). Probe mixing experiments, using a *Ubx* probe as a parasegment (PS)²⁷ marker, reveal that the most strongly labelled neuromeres correspond to PS3, 4, 5 and 14. This is strikingly complementary to the pattern of *Ubx* expression, which extends from PS5–13, with the expression in PS5 being relatively weak^{28,29}. In the case of clone 48, transcripts initially accumulate in a ventral stripe of the blastoderm (stage 5) with no relation to *Ubx* expression (Fig. 4b). After germ-band extension (stage 10), a clear pattern of segmentally repeated labelling is apparent in some epidermal cells of the anterior compartments of the first thoracic to the ninth abdominal segments (T1–A9; PS3–14). From stage 12, transcripts seem restricted to ventrolateral and dorsolateral subsets of each anterior compartment (Fig. 4c, d) with clear segment-to-segment modulation in the labelling. The ventrolateral labelling is considerably reduced in T2 (PS4) and T3 (PS5) relative to other labelled segments. The differences between PS4, 5 and 6 are highly suggestive of regulation by *Ubx*.

The 9 and 72 transcripts also show segment-to-segment modulation but this is not so obviously correlated with the *Ubx* distribution. The 9 transcripts are expressed at high levels in PS15 and the 72 transcripts accumulate in the mesoderm of a subset of abdominal segments (data not shown).

Transcript regulation by *Ubx*

To investigate whether the *Ubx* locus controls the observed segment-to-segment modulations in the expression patterns of the 35 and 48 transcripts, transcript distributions were examined in embryos deficient for the *Ubx* gene (Fig. 5). These embryos show transformations of PS5 and PS6 towards PS4 with minor transformations more posteriorly^{2,30–33}.

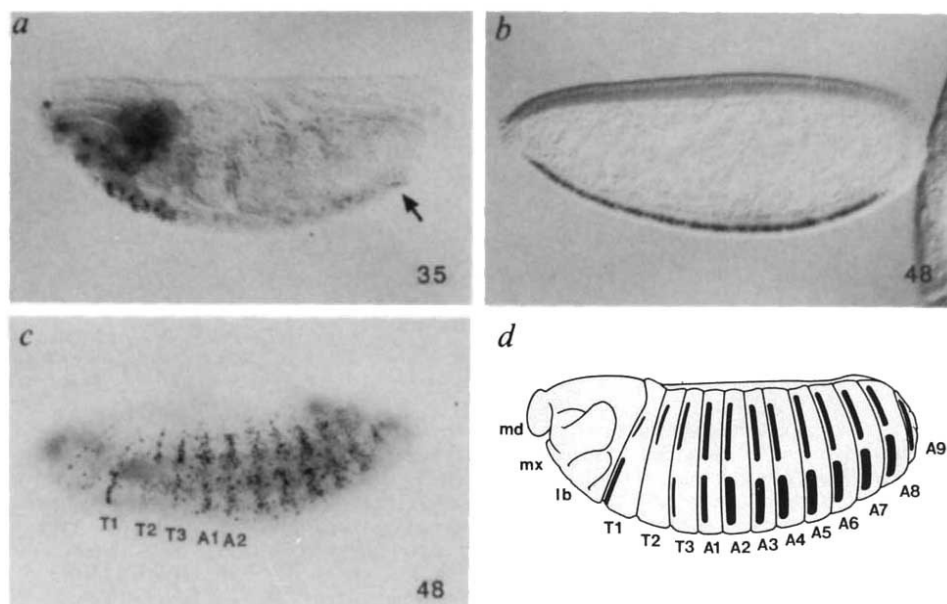
The expression of the 35 transcripts was studied in CNS preparations dissected from late embryos, enabling all the parasegments to be visualized within one focal plane. The down-regulation that, in the wild type, occurs between PS6–13 (Fig. 5a), is observed only in PS7–13 in *Ubx* mutant embryos (Fig. 5b). This demonstrates that *Ubx* functions to repress levels of 35 transcript in PS6 of the wild type. Repression of the 35 transcripts shows a threshold response with the high levels of *Ubx* protein in PS6 producing strong repression, whereas the low levels in PS5 seem to have little effect. The repression in PS7–13 seen in *Ubx* mutant embryos is eliminated in embryos deficient for the entire bithorax complex (that is, *Ubx*[−], *abdominal-A*[−], *Abdominal-B*[−], ref. 33; Fig. 5c) and thus is mediated by *abdominal-A* (expressed from PS7–13)³⁴ with the possibility of some contribution from *Abdominal-B* (expressed in PS10–14)³⁵.

The 48 transcript distribution was examined in *Ubx* mutants in the ventrolateral epidermis of T1–A2 of the shortened germ band (stage 14). In wild type embryos, the 48 transcripts are high in T1 (PS3), barely detectable in T2 (PS4), but at successively higher levels in T3 (PS5) and A1 (PS6), reaching a maximum in A2 (PS7) and more posterior abdominal segments (Fig. 5d). In stage 14 embryos deficient for the *Ubx* locus, a reduction is consistently observed in the mutant T3 (T3*), and is particularly clear in the mutant A1 (A1*). Thus, in these mutant embryos, the 48 transcript levels in T3* and A1* are similar to the low levels found in T2 of the wild type, whereas levels in T1 and A2–9 are largely unaffected (Fig. 5e). Hence, the *Ubx* locus functions to positively regulate the 48 transcript levels in T3–A1 (PS5–6) of the wild type ventrolateral epidermis.

In vivo targets of *Ubx*

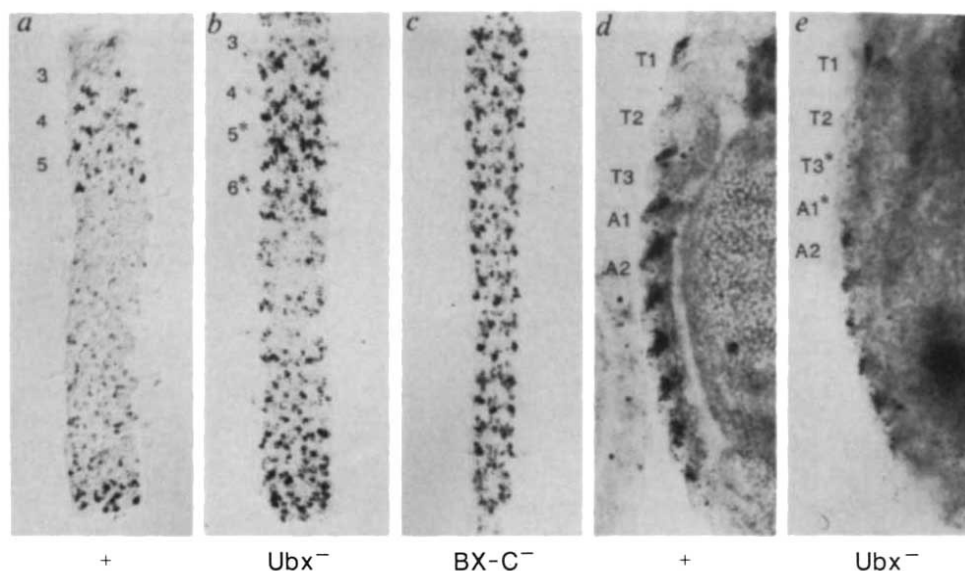
We have described an immunopurification strategy, using embryonic chromatin as a substrate, that has yielded at least two good candidates for target genes regulated by the homeotic gene *Ubx*. Other immunopurification approaches have been described previously^{36–39}. Most of these have attempted to reconstitute, *in vitro*, relevant interactions between a purified DNA-binding protein and genomic DNA^{37–39}. However, many

FIG. 4 Wild-type distribution of the 35 and 48 transcripts. Embryos are shown with anterior to left and staged as described²⁶. Transcript numbers are indicated in bottom right corner of each photomicrograph. *a*, Lateral view of a stage 16 embryo hybridized with a 10 kb *SalI* fragment from the 35 region. Transcripts are most abundant in the brain lobes (out of focus), three thoracic neuromeres and the most posterior abdominal neuromere (arrowed, see also Fig. 5). *b* and *c*, Lateral view of stage 5 and stage 14 embryos, respectively. Both embryos were hybridized with a 7.4 kb *SalI* fragment from the 48 region. At stage 5, transcripts are present in a ventral region including the presumptive mesoderm. In the stage 14 embryo, transcripts are detected in dorsolateral and ventrolateral epidermal cells of the anterior compartments of T1–A9 (by reference to the segmental grooves). Within the T1–A2 region (indicated in *c*), there are clear segmental differences in transcript levels (see legend to Fig. 5). *d*, Schematic representation of *c* shows mandibular (md), maxillary (mx), labial (lb), thoracic and abdominal segments with regions expressing high levels of the 48 transcripts indicated. METHODS. Inserts from +antibody clones 35 and 48 were used as probes to isolate λ phage, at high stringency, from an EMBL3 genomic library (gift



of J. Tamkun) using standard methods⁴⁸. *SalI* insert fragments from these phage were chemically labelled and used as probes in whole mount *in situ* hybridizations to 0–24 h Oregon R embryos⁴⁹.

FIG. 5 Distribution of transcripts from the 35 and 48 regions in both wild type and homeotic mutant embryos. Anterior is uppermost and the approximate positions of parasegments (a–c) and segments (d, e) are indicated. Asterisks denote major units affected in *Ubx* mutants. a–c, The 35 transcripts in ventral nerve cords of late wild type (+), *Ubx* deficient (*Ubx*[−]), and bithorax complex (BX-C) deficient (BX-C[−]) embryos. The approximate position of the PS5/6 (a), or PS6*/7 (b) border is indicated. In the wild type, highest levels of transcripts are in cell clusters in PS3–5 and in a posterior domain including PS14. Note that this posterior expression may indicate lack of repression by the *Abdominal-B* protein^{50,51}. *Ubx*[−] nerve cords show a clear derepression of transcript levels in PS6* and more subtle but consistent derepressions in PS7–13. BX-C[−] nerve cords show uniform high levels of the 35 transcripts. d, e, The 48 transcripts in the ventrolateral epidermis of stage 14 wild type (+) and *Ubx* deficient (*Ubx*[−]) embryos, respectively. In the wild type, transcripts are at low levels in T2 but at successively higher levels in T3 and A1, reaching a maximum in A2 and more posterior segments. In *Ubx*[−] embryos, transcript levels in T3* and A1* become reduced towards the low levels normally found only in T2.



legend to Fig. 5. Embryos (0–24 h) were derived from Oregon R, Df(3R)*bxd*¹⁰⁰/TM1, or Dp(3:3)P5/Df(3R)P9 fly stocks. ~25% of embryos from Df(3R)*bxd*¹⁰⁰/TM1 or Dp(3:3)P5/Df(3R)P9 stocks are deficient for the *Ubx* gene or the entire BX-C, respectively^{2,33}. CNS preparations were dissected, on coverslips coated with poly-L-lysine, from embryos after the alkaline-phosphatase colour reaction.

of the target sites isolated using such *in vitro* strategies may not be occupied *in vivo* because of constraints on binding imposed by chromatin structure^{40–42} or the presence of other competing factors⁴³.

The total number of *Ubx* target sites in a nucleus is unknown, and it is unclear what fraction of these *in vivo* target sites would be expected to mediate homeotic gene control. It is probable that the immunopurification method selects for target sites capable of a stable interaction with *Ubx* protein. This may bias the immunopurified population in favour of functional binding sites.

The two transcripts that we have investigated in detail respond differently to *Ubx* regulation. Expression of the 35 transcripts is down-regulated, whereas the 48 transcripts are positively regulated. This suggests that *Ubx* may function either to activate or to repress target loci, depending on the exact cellular and promoter contexts. This is consistent with the dual roles of *Ubx* in positive autoregulation¹⁶ and in repression of *Antennapedia*^{14,44}.

Neither the 35 nor the 48 transcript distributions is correlated with the *Ubx* expression pattern alone. Rather, the patterns indicate regulation by several homeotic genes. In particular, the derepression of the 35 transcripts observed in PS7–13 of embryos deficient for the bithorax complex, indicates an additional negative regulation by at least the *abdominal-A* gene. That several homeotic genes may regulate the same target gene is consistent with the similarities in DNA-binding specificities reported for many homeotic gene products. Also, studies on the transformations induced by the ectopic expression of homeotic genes suggest a large overlap between their target gene pools⁴⁵.

The morphogenetic functions that are mediated by the subordinate targets of homeotic genes should include such region-specific cell properties as mitotic rate, mitotic orientation, cell-surface adhesion and cuticular differentiation³. The transcript distributions that we have described give few clues to the functions of the genes, however, the dynamic pattern of the 48 transcripts is interesting as it suggests a role in processes common to both early dorsoventral patterning and late epidermal differentiation. The expression patterns of the 35 and 48 transcripts do not suggest a correspondence with any previously

cloned genes within the cytological intervals, 64C and 97D, respectively.

In summary, we present the genes encoding the 35 and 48 transcripts as good candidates for target genes directly regulated by the homeotic gene *Ubx*. Although the genetic evidence for *Ubx*-dependent regulation of these transcripts does not distinguish between direct and indirect interactions, we believe that the regulation is direct for two reasons; first, we have identified these two transcription units only because of their close proximity to chromatin fragments immunopurified with *Ubx* protein and second, both immunopurified fragments contain specific binding sites for *Ubx* protein *in vitro*. We are investigating whether these *Ubx* protein-binding sites are required *in vivo* to mediate the observed homeotic control of the 35 and 48 transcripts. □

Received 15 August; accepted 18 October 1990.

- Lewis, E. B. *Am. Zool.* **3**, 33–56 (1963).
- Lewis, E. B. *Nature* **276**, 565–570 (1978).
- Garcia-Bellido, A. in *Ciba Fdn Symp.* **29**, 161–178 (1975).
- Gehring, W. J. *Science* **236**, 1245–1252 (1987).
- Scott, M. P., Tamkun, J. W. & Hartzell, G. W. *BBA Rev. Cancer* **989**, 25–48 (1989).
- Desplan, C., Theis, J. & O'Farrell, P. H. *Nature* **318**, 630–635 (1985).
- Beachy, P. A., Krasnow, M. A., Gavis, E. R. & Hogness, D. M. *Cell* **55**, 1069–1081 (1988).
- Müller, M. *et al.* *EMBO J.* **7**, 4299–4304 (1988).
- Thali, M., Müller, M. M., Delorenzi, M., Matthias, P. & Bienz, M. *Nature* **336**, 598–601 (1988).
- Krasnow, M. A., Saffman, E. E., Kornfeld, K. & Hogness, D. S. *Cell* **57**, 1031–1043 (1989).
- Winslow, G. M., Hayashi, S., Krasnow, M. A., Hogness, D. S. & Scott, M. P. *Cell* **57**, 1017–1030 (1989).
- Samson, M.-L., Jackson-Grusby, L. & Brent, R. *Cell* **57**, 1045–1052 (1989).
- Johnson, F. & Krasnow, M. A. *Genes Dev.* **4**, 1044–1052 (1990).
- Hafen, E., Levine, M. & Gehring, W. J. *Nature* **307**, 287–289 (1984).
- Struhl, G. & White, R. A. H. *Cell* **43**, 507–519 (1985).
- Bienz, M. & Tremml, G. *Nature* **333**, 576–578 (1988).
- Kuziora, M. A. & McGinnis, W. *Cell* **55**, 477–485 (1988).
- Laughon, A., Howell, W. & Scott, M. P. *Development (suppl.)* **104**, 75–83 (1988).
- Goutte, C. & Johnson, A. D. *Cell* **52**, 875–882 (1988).
- Keleher, C. A., Goutte, C. & Johnson, A. D. *Cell* **53**, 927–936 (1988).
- Gould, A. P. thesis, Univ. Cambridge, UK (1990).
- White, R. A. H. & Wilcox, M. *Cell* **39**, 163–171 (1984).
- Noll, M., Thomas, J. O. & Kornberg, R. D. *Science* **187**, 1203–1206 (1975).
- Desplan, C., Theis, J. & O'Farrell, P. H. *Cell* **54**, 1081–1090 (1988).
- Staden, R. *Nucleic Acid Res.* **12**, 505–519 (1984).
- Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1985).
- Martinez-Arias, A. & Lawrence, P. A. *Nature* **313**, 639–642 (1985).
- Akam, M. & Martinez-Arias, A. *EMBO J.* **4**, 1689–1700 (1985).
- White, R. A. H. & Wilcox, M. *EMBO J.* **4**, 2035–2043 (1985).

30. Morata, G. & Kerridge, S. *Nature* **290**, 778–781 (1981).
31. Hayes, P., Sato, T. & Denell, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 545–549 (1984).
32. Struhl, G. *Nature* **308**, 454–457 (1984).
33. Sánchez-Herrero, E., Vernós, I., Marco, R. & Morata, G. *Nature* **313**, 108–113 (1985).
34. Harding, K., Wedeen, C., McGinnis, W. & Levine, M. *Science* **229**, 1236–1242 (1985).
35. Celniker, S. E., Keelan, D. J. & Lewis, E. B. *Genes Dev.* **3**, 1424–1436 (1989).
36. Lane, D. P. *et al. Proc. R. Soc. B226*, 25–42 (1985).
37. Biedenkapp, H., Borgmeyer, U., Sippel, A. & Klempnauer, K.-H. *Nature* **335**, 835–837 (1988).
38. Gruss, C., Wetzel, E., Baak, M., Mock, U. & Knippers, R. *Virology* **167**, 349–360 (1988).
39. Sompayrac, L. & Danna, K. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3274–3278 (1990).
40. Almer, A., Rudolph, H., Hinnen, A. & Hörz, W. *EMBO J.* **5**, 2689–2696 (1986).
41. Thomas, G. H. & Elgin, S. C. R. *EMBO J.* **7**, 2191–2201 (1988).
42. Pina, B., Brüggermeier, U. & Beato, M. *Cell* **60**, 719–731 (1990).
43. Vincent, J.-P., Kassiss, J. & O'Farrell, P. *EMBO J.* **9**, 2573–2578 (1990).
44. Carroll, S. B., Laymon, R. A., McCutcheon, M. A., Riley, P. D. & Scott, M. P. *Cell* **47**, 113–122 (1986).
45. Gibson, G. & Gehring, W. J. *Development* **102**, 657–675 (1988).
46. Wu, C. *Nature* **309**, 229–234 (1984).
47. Langer-Safer, P. R., Levine, M. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4381–4385 (1982).
48. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning, A Laboratory Manual*, 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
49. Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81–85 (1989).
50. Casanova, J., Sánchez-Herrero, E. & Morata, G. *Cell* **47**, 627–636 (1986).
51. Delorenzi, M. *et al. EMBO J.* **7**, 3223–3231 (1988).

ACKNOWLEDGEMENTS. We thank D. Tautz and C. Pfeifle for communicating the whole-mount *in situ* hybridization method before publication, M. Wilcox and R. Smith for help in the early stages of this project, and M. Ashburner for advice on chromosome localizations. This work was supported by the MRC and Wellcome Trust.

LETTERS TO NATURE

The extended sodium nebula of Jupiter

Michael Mendillo, Jeffrey Baumgardner,
Brian Flynn & W. Jeffrey Hughes

Department of Astronomy and Center for Space Physics,
Boston University, Boston, Massachusetts 02215, USA

THE detection of a cloud of neutral sodium near Jupiter's moon Io¹ has led to the use of sodium as a tracer of processes in the jovian environment. Although relatively rare in the Io–Jupiter system, sodium atoms are easily detected because of their high efficiency for scattering sunlight at wavelengths of $\sim 5,890$ Å. Direct imaging of the sodium cloud² has suggested that sodium atoms are a common feature close to Io (at distances of about six Io radii, R_{Io}) and detection of high-speed sodium jets³ suggested that sodium is present only sporadically at $\sim 30R_{\text{Io}}$ (ref. 4). Sodium emission has been reported at greater distances⁵, even as far as $60R_{\text{Io}}$ (ref. 6) but these observations have been controversial in view of suggestions⁷ that the detection of sodium beyond $\sim 10R_{\text{Io}}$ was implausible on theoretical grounds and probably indistinguishable from terrestrial sodium airglow. Here we report on the detection of sodium to distances beyond $\sim 400R_{\text{J}}$, an observation that requires the ejection rate of sodium atoms to be increased. By relating the shape of this great nebula to conditions in the plasma torus surrounding Jupiter, we show that ground-based imaging techniques can provide information about distant planetary magnetospheres.

An observational programme was initiated using a low-dispersion high-sensitivity imaging spectrograph⁸ to search for emissions from ionic and neutral species at the great distances from Jupiter that would be appropriate for imaging magnetospheric structures. Our observations were made at the McDonald Observatory in Fort Davis, Texas, during the period 26 November–5 December 1989. Sodium emission was by far the strongest signal of jovian origin obtained at distances from 50 to $300R_{\text{J}}$, and a subsequent period of sodium-imaging observations was therefore conducted from 20 to 30 January 1990 using our standard intensified CCD detector⁹ and a portable 4.2-inch aperture telescope. We obtained sodium images of 320-s exposure using a filter with a transmission of 51% and a full width at half power (FWHP) of 12 Å centred at $5,893$ Å, thereby including both the D_1 ($5,896$ Å) and D_2 ($5,890$ Å) lines. Control images were taken using a filter with 60% transmission centred at $6,200$ Å with a FWHP of 14 Å. Images were taken with the instrument's 6° field of view centred on Jupiter and at a point 10° north of Jupiter. We used an occulting shield and therefore did not obtain usable data in the region $\pm 25R_{\text{J}}$ E–W and $\pm 13R_{\text{J}}$ N–S.

Figure 1 gives raw data sets taken on 25 January 1990. In Fig. 1a, with the $5,893$ Å filter on Jupiter, light contributions come from sodium in Jupiter's environment, Jupiter's light scattered by the Earth's atmosphere and within the instrument, emissions from sodium in the Earth's atmosphere, and emission from stars

and terrestrial OH bands that fall in the filter's transmission curve. In Fig. 1b, taken with the control filter on Jupiter, light contributions come from Jupiter's scattered light, the same star field and terrestrial OH. In Fig. 1c, d, identical type images taken 10° north of Jupiter contain all of the non-jovian emissions (and a different star field).

After correcting each of the four images for instrumental effects (vignetting, bias and gain), we constructed the final image from $[(\text{Na} - \text{control})_{\text{Jupiter}} - (\text{Na} - \text{control})_{\text{Off-Jupiter}}]$, that is, (Fig. 1a – Fig. 1b) – (Fig. 1c – Fig. 1d). In addition, both of the control images were scaled to account for differences between the two filters, sensitivity of the instrument and the spectral distribution of scattered light. The resultant image, corrected for atmospheric extinction (20%) and the angular dependence of the sodium filter, is shown in Fig. 1e. We determined intensities using a ^{14}C source—a 3-inch diameter phosphor, illuminated uniformly by radioactively decaying carbon, that emits $80\text{ R } \text{Å}^{-1}$ at $5,893$ Å (calibrated against a tungsten halogen lamp certified by the US National Bureau of Standards).

Figure 2 shows the intensity distribution along the main axis of the nebula in Fig. 1e, together with an inverse-distance intensity model. Moving outward from Jupiter, the brightness levels are clearly not symmetrical, with stronger signals extending to $100R_{\text{J}}$ on the westward side (where Io was). Intensity levels beyond $100R_{\text{J}}$ (approximately the location of Jupiter's magnetopause) are fairly symmetrical out to $\sim 400R_{\text{J}}$.

Figure 3 shows the north–south structure of the nebula at several radial distances. The coning or flaring angle in the pattern has a width of $\sim \pm 22^\circ$ at half-intensity. This angle is greater than the 7° tilt between Io's orbital plane (where the sodium cloud resides) and Jupiter's centrifugal equator (where the Io plasma torus densities are at their maximum). The shape of the nebula is thus not so much determined by the geometry of the interaction between the torus and the sodium cloud as by the plasma conditions and dynamics at Io's orbit, as discussed below.

We note that, as D_1 intensities are $\sim 50\%$ those of D_2 , our intensities ($D_1 + D_2$) are a factor of ~ 1.5 greater than the D_2 -only intensities most often quoted. The equivalent D_2 intensities from our innermost observations (50 – $70R_{\text{J}}$ at $25R_{\text{J}}$) in Figs 1e and 2 are factors of two to three higher than those in some earlier reports of remote sodium: ~ 15 – $30R_{\text{J}}$ (D_2) at 15 – $20R_{\text{J}}$ (ref. 10) and ~ 25 – $30R_{\text{J}}$ (D_2) at $35R_{\text{J}}$ (ref. 5), although basically consistent with others $\sim 30R_{\text{J}}$ (D_2) at $60R_{\text{J}}$ (ref. 6 and L. Trafton, personal communication) and $\sim 50R_{\text{J}}$ (assumed D_2) at 15 – $25R_{\text{J}}$ (ref. 11). The vast extent of the nebula is revealed in the very faint ($\leq 5\text{ R}$) levels which extend to the edge of our field of view. Our calibration procedures were developed specifically for low surface brightness features and were checked as follows: terrestrial sodium levels of $\sim 24\text{ R}$ (D_1 and D_2) given in Fig. 1c are comparable to published values¹²; use of our imaging system for H_α studies of a portion of Barnard's loop during the same January 1990 period yielded intensities in agreement with published values¹³.

Owing to the strong sodium absorption feature in the solar spectrum, the relationship between intensity levels and sodium

column content along the line of sight depends on both the density and velocity structure of the nebula. Modelling studies currently underway, similar to those of other groups^{10,14}, suggest that fast sodium atoms ejected from the cloud near Io move along escape trajectories that are approximately tangential to co-rotation near Io ($V_r = 74 \text{ km s}^{-1}$ at $\sim 6 R_J$), leading eventually to a radially streaming gas at great distances ($\geq 50 R_J$). Assuming minimal sources and sinks beyond Io's orbit, and a constant radial velocity, yields a r^{-2} density structure. For low radial velocities ($V_r \approx 0$), the intensities in Fig. 1e are related to column content by $\epsilon_r = 10^{-6} g N_T$, where ϵ_r is the emission in Rayleigh units at distance r in the equatorial plane, g is 0.03 photons ($D_1 + D_2$) scattered per Na atom per second, and N_T is the column content in atom cm^{-2} along the line of sight. Thus, at $100 R_J$, $\epsilon_r = 23 \text{ R}$ corresponds to $7.5 \times 10^8 \text{ Na cm}^{-2}$. With a substantial radial velocity with respect to Jupiter (and thus to the Sun), $V_r \approx 74 \text{ km s}^{-1}$, our preliminary modelling studies of the resultant spectral-line profiles suggest that the measured ϵ_r represents a factor of 18 enhancement over the $V_r \approx 0$ case, yielding a column content of $4.3 \times 10^7 \text{ Na cm}^{-2}$. Modelling the sodium distribution along this column gives a concentration ($[\text{Na}] =$

$N_T / \pi r$) of $\sim 2 \times 10^{-5} \text{ cm}^{-3}$ at $r = 100 R_J$. The flux required to maintain such a distribution with a flaring angle of $\pm \theta$ is $\Phi_{\text{Na}} = [\text{Na}] V_r 2\pi r (2\theta) \approx 4 \times 10^{26} \text{ Na s}^{-1}$. Following the Brown and Schneider formalism¹⁰ for relating the flux of fast sodium from charge exchange to the total inventory of sodium near Io, we derive a total abundance value of $\sim 8 \times 10^{31}$ atom, a factor of 80 higher than used in that model.

Our analysis thus far shows: (1) the spatial extent of sodium emission far from Jupiter is a factor of 10 larger than previously known; (2) total intensities ($D_1 + D_2$) are consistent with some earlier D_2 intensities observed in the 25–50 R_J region, yet at least a factor of two higher than others, suggesting that temporal variability may be an observable component of the nebula; (3) the fast-sodium source rate calculated from observations at $100 R_J$, when interpreted in terms of standard cross-sections and model parameters¹⁰, requires an increase in the total sodium population, or the Na^+ densities in the torus, or both (such that both multiplicative factors result in an increase by a factor of ~ 80). The most recent estimates of fluxes associated with fast sodium jets¹⁵ are also well below (by a factor of 50) those we obtain for the nebula.

Perhaps the most surprising aspect of these observations is the well ordered shape of the sodium nebula shown in Fig. 1e and characterized in Fig. 3 by a flaring angle of $\sim \pm 22^\circ$. We relate this angle to the co-rotation speed and temperature of ions in the plasma torus. Assuming that resonant charge exchange ($\text{Na}^+ + \text{Na} \rightarrow \text{Na}(\text{fast}) + \text{Na}^+$) represents the main avenue for the production of fast sodium¹⁰, the flaring angle serves as a diagnostic of the Na^+ mean energy in the torus. If the velocity-space distribution of ions in the torus is assumed to be spherical about the co-rotation speed of $\sim 74 \text{ km s}^{-1}$, the radius of the sphere is $74 \sin 22^\circ$ or $\sim 28 \text{ km s}^{-1}$. This velocity corresponds to a sodium ion of energy 94 eV. Although Na^+ is a minor species in the torus ($< 5\%$), its thermal structure should be a tracer for the total ion population. The derived value of 94 eV is consistent with the mean ion energy of $\sim 100 \text{ eV}$ in the torus at $6 R_J$ reported from the 1979 Voyager encounters¹⁶. The strong control exerted by such magnetospheric processes on the sodium distribution far from its source results in a structure decidedly different from the gravitationally bound cloud and occasional jets of sodium long associated with Io's extended atmosphere. The fact that the shape of the structure can be used to conduct remote sensing of Jupiter's magnetospheric properties suggest a new term for such a non-spherical planetary neutral cloud, a magneto-nebula.

The ultimate fate of the sodium atoms ejected from Jupiter's magnetosphere seems to be ionization by solar extreme ultraviolet with a time constant of $\sim 400 \text{ h}$ (ref. 4). At $\sim 74 \text{ km s}^{-1}$ (or $\sim 4 R_J \text{ h}^{-1}$), the nebula would extend to $\sim 1.5 \text{ AU}$ in diameter

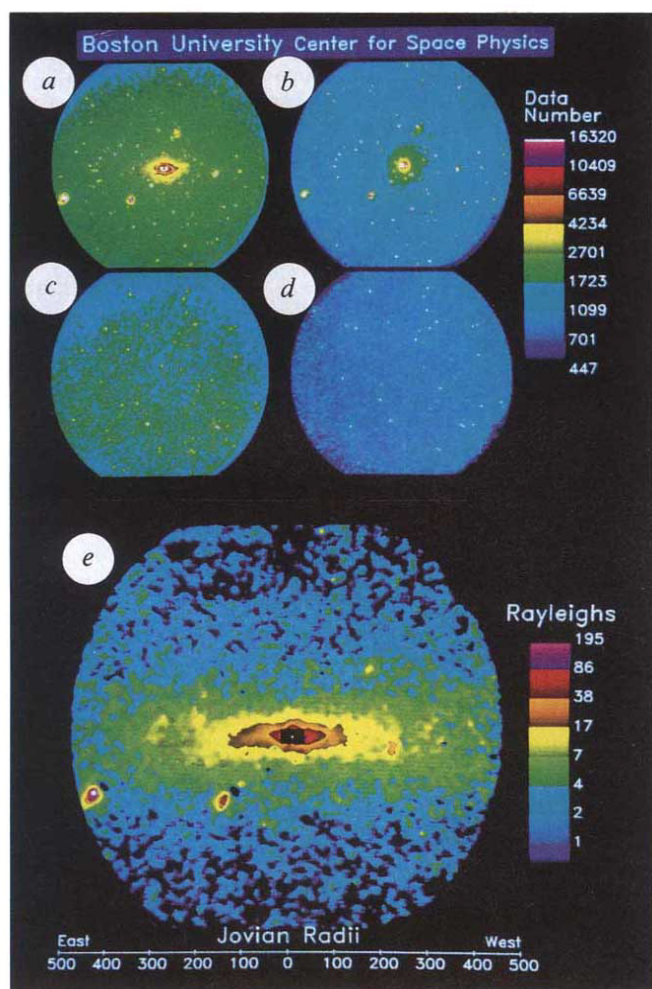


FIG. 1 Images taken on 25 January 1990 with a, sodium ($5,893 \text{ \AA}$) filter at 02:54:09 UT and b, control ($6,200 \text{ \AA}$) filter at 02:25:40 UT, both centred on Jupiter, and c, at 02:44:54 UT and d, at 02:36:43 UT with the sodium and control filters 10° north of Jupiter. A radially symmetrical pattern of scattered light is given in b. The terrestrial sodium brightness level from c was found to be $\sim 24 \text{ R}$. The background through the $6,200\text{-\AA}$ filter from d was $\sim 1.5 \text{ R } \text{\AA}^{-1}$. The brightness levels in a–d are given in original data numbers to allow for comparison. The final corrected image is shown in e (see text). The two bright features in the southeast quadrant are residuals from the incomplete subtraction of stars μ and η Geminorum.

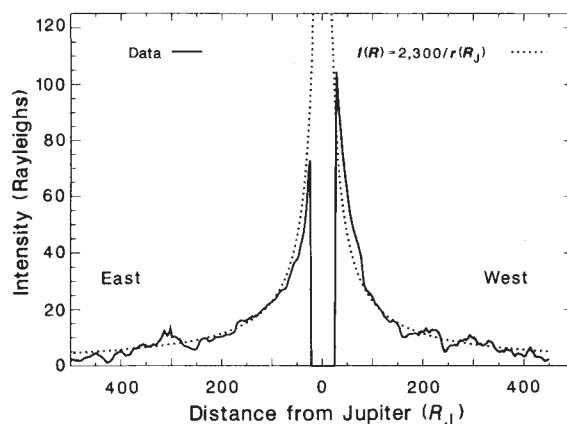


FIG. 2 Scan of the intensities in Fig. 1e taken along a $9\text{-}R_J$ -wide (5-pixel-wide) band in the equatorial plane. For comparison, an inverse-distance brightness profile is shown that is normalized to the observed values of 23 R at $\pm 100 R_J$.

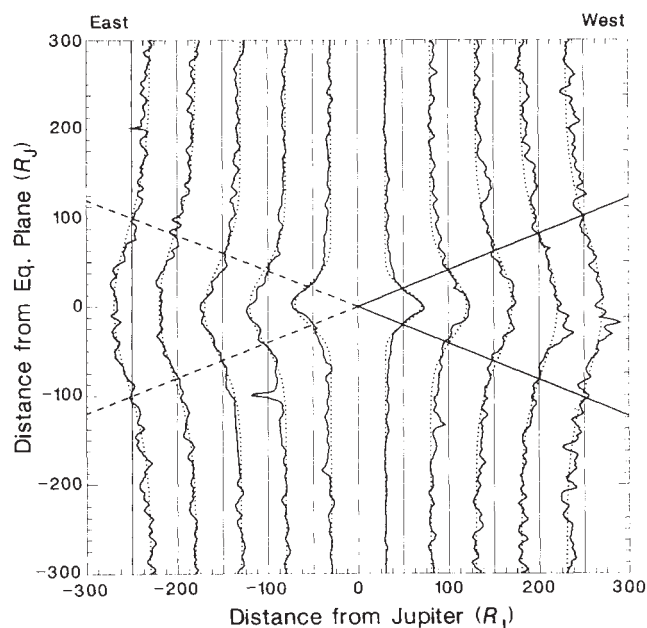


FIG. 3 Scans perpendicular to the equatorial plane at several radial distances from Jupiter. The scans have been normalized to unit intensity at the equatorial crossing point to illustrate the flaring angle defined by the half width at half maxima obtained from gaussian fits to the data (dotted curves). The western and eastern portions of the image were processed separately and resulted in nearly identical flaring angles: 22° (solid lines) and 21.3° (dashed lines). The two westernmost scans are contaminated by imperfect subtraction of light from 1 Geminorum. The small tilt of Jupiter's equatorial plane ($\sim 2^\circ$) with respect to Earth was not removed before this analysis because it has no appreciable effect on the measured half-widths.

before substantial ionization occurred. This faint cloud is perhaps the largest visible object in the Solar System, and its full extent might be detected by a space-based imaging system above the terrestrial sodium layer and scattering atmosphere. The Galileo spacecraft might also detect neutral sodium as it approaches Jupiter in 1995. Such energetic-neutral-atom imaging of magnetospheric topology has been discussed for the terrestrial and saturnian cases¹⁷.

The mechanisms that produce Jupiter's sodium magnetonebula should also create extended clouds of the more abundant species (such as O, S, SO, SO₂). They have far weaker signatures at visible wavelengths, but might appear in space-borne ultraviolet images (the emission at 1,304 Å of O and S, for example) or in direct detections of neutral fluxes by the Galileo spacecraft. □

Received 22 May; accepted 16 October 1990.

1. Brown, R. A. in *Proc. IAU Symp. 65, Exploration of the Planetary Systems* (eds Woszczyk, A. & Iwaniszewska, C.) 527–531 (Reidel, Dordrecht, 1974).
2. Goldberg, B. A., Garneau, G. W. & LaVoie, S. K. *Science* **226**, 512–516 (1984).
3. Trauger, J. T. *Science* **226**, 337–341 (1984).
4. Schneider, N. M., Smyth, W. H. & McGrath, M. A. *Time-Variable Phenomena in the Jovian System*, SP-494 (eds Belton, M. J. S., West, R. A. & Jurgens, R.) 75–94 (NASA, Washington, 1989).
5. Pilcher, C. B. & Schempff, W. V. *Icarus* **38**, 1–11 (1979).
6. Trafton, L. & Macy, W. Jr. *Icarus* **33**, 322–325 (1978).
7. Goody, R. & Aft, J. *Planet. Space Sci.* **25**, 603–604 (1977).
8. Baumgardner, J. *Eos* **70**, 404 (1989).
9. Baumgardner, J. & Karandanis, S. *Electronic Imaging* **3**, 28–31 (1984).
10. Brown, R. A. & Schneider, N. M. *Icarus* **48**, 519–535 (1981).
11. Wehinger, P. A., Wyckoff, S. & Frohlich, A. *Icarus* **27**, 425–528 (1976).
12. Kirchhoff, V. W. J. H. *Can. J. Phys.* **64**, 1664–1672 (1986).
13. Reynolds, R. J. & Ogden, P. M. *Astrophys. J.* **229**, 942–953 (1979).
14. Pilcher, C. B., Smyth, W. H., Combi, M. R. & Fertel, J. H. *Astrophys. J.* **287**, 427–444 (1984).
15. Schneider, N. thesis, Univ. Arizona (1988).
16. Bagenal, F. in *Time-Variable Phenomena in the Jovian System*, SP-494 (eds Belton, M. J. S., West, R. A. & Jurgens, R.) 196–210 (NASA, Washington, 1989).
17. Roelof, E. C. & Williams, D. J. *Johns Hopkins APL Tech. Digest* **9**, 72–76 (1990).

ACKNOWLEDGEMENTS. We thank A. Dessler, F. Bagenal, N. Schneider and G. Siscoe for discussions, E. Barker and the McDonald Observatory for support, K. Jones, R. Phelps and M. Fox for help with planning and analysis, D. Nottingham and H. Pais for assistance in the observations, S. Karandanis (Datacube Inc.) for providing digitizing hardware, and T. Mooney (Barr Assoc.) for a sodium filter. We dedicate this paper to the memory of F. Scarf.

Evidence of very-large-amplitude solitary waves in the atmosphere

Mohan K. Ramamurthy*, Brian P. Collins*, Robert M. Rauber* & Patrick C. Kennedy†

* Department of Atmospheric Sciences, University of Illinois, 105 S Gregory, Urbana, Illinois 61801, USA

† Climate and Meteorology Section, Illinois State Water Survey, Champaign, Illinois 61820, USA

ATMOSPHERIC solitary waves are gravity waves that retain their integrity over long periods because of a near balance between nonlinearity and dispersion. They have been observed on various scales in many regions of the world^{1–3}, but we present here detailed measurements of solitary waves with amplitudes comparable to the scale height of the lower troposphere. Two such waves were generated downstream of intense mid-tropospheric pressure troughs over the central United States. They propagated over 1,000 km (several times their wavelength) with no appreciable change in structure within a 'waveguide' formed by surface inversion and a middle tropospheric critical level. Fluctuations in surface pressure associated with the two waves exceeded 6 mbar and 10 mbar. The waves caused banded patterns of precipitation and significantly influenced other meteorological phenomena. The restoration of balance between pressure-driven air flow and the Coriolis force ('geostrophic adjustment') seems to have a prominent role in the formation of these solitary waves.

Studies have documented the existence of atmospheric solitary waves during the winter^{4–7} and have revealed a large degree of similarity in the atmospheric conditions present during their initiation⁸. Several mechanisms have been proposed and documented to explain their genesis, including the hydraulic jump theory^{9,10}, a shearing instability mechanism^{6,11,12}, the geostrophic adjustment process^{6,8,13}, interactions of density currents with pre-existing inversions^{10–14}, and triggering by deep convection². Although several hypotheses have been forwarded, the lack of detailed measurements through the depth of the troposphere has limited our understanding of these phenomena.

During January 1989, two large-amplitude solitary waves passed through the central United States. Throughout both events, thermodynamic and wind data were gathered at high time resolution with three instruments: a 50-MHz wind profiler¹⁵, a 10-cm-wavelength Doppler radar¹⁶ and a balloon-borne sounding system¹⁷. These measurements were made near Champaign, Illinois during an investigation of mid-latitude cyclone precipitation band structure.

The solitary waves of 5 and 14 January 1989 developed within intense wintertime mid-latitude cyclones. The waves in both cases propagated over 1,000 km before dissipating (Fig. 1). Dramatic surface pressure fluctuations were recorded at many stations. For example, the pressure dropped 10 mbar in 75 min at Cape Girardeau on 14 January, and 6 mbar over the same time period at Moline on 5 January. The arrival of the wave at many locations, as evidenced by coherent surface pressure jumps in Fig. 1, was marked by gustiness and sharp changes in the wind direction and intensity of precipitation. Narrow, very distinct precipitation bands accompanied the waves on both days (Fig. 2). The return to equilibrium pressures near pre-wave levels following wave passage (after subtracting the larger-time-scale trend) suggests that these disturbances possessed the characteristics of solitary waves rather than undular bores, where a finite pressure jump would be expected¹⁸.

In both cases, deep pressure troughs in the middle and upper troposphere were located west of the region of solitary wave generation. In each trough, strong winds at jet-stream altitude moved northeast from the trough base into a region of reduced pressure gradient and weak winds. In such a flow pattern the ratio of the air-parcel acceleration to the Coriolis force is large,

resulting in an unbalanced state. The adjustment back to a balanced state, termed geostrophic adjustment¹⁹, redistributes the mass and momentum fields, and has been suggested as a potential mechanism for the generation of solitary waves⁸. Geostrophic adjustment should trigger inertia gravity waves that induce large downward displacements of upper tropospheric air in the deceleration region downwind of the jet-stream maximum.

Atmospheric solitary waves have been identified previously as either waves of depression or elevation^{20,21}. The wave-normal streamlines and the wave-normal wind speed (u) time series in Fig. 3 show the presence of a large-amplitude solitary wave of elevation over Champaign between 22:00 and 00:00 UT. The passage of the solitary wave, manifested as a precipitation band over Champaign (Fig. 2), coincided with the time of maximum pressure drop at the surface, and the strongest upward vertical motions. Rapid clearing of radar echoes and strong gusts in the surface winds accompanied the pressure rise as the wave moved east of the radar site. These observations, and the analysis in

Fig. 3, suggest strong wave-induced subsidence and a strong downward-directed momentum flux on the west side of the wave crest. The 5 January time series of u (Fig. 3) also shows evidence of trapping and subsequent wave reflection. There is an apparent phase shift during wave passage, changing from an upstream tilt ahead of the wave, to a downstream tilt after wave passage, evidence of wave reflection at the 6-km level.

In our 5 January observations, we found that the same disturbance produced an initial positive surface pressure perturbation at Moline, but a negative pressure perturbation at Springfield, Champaign and Indianapolis, (Fig. 1). Despite the negative pressure perturbation, Fig. 3 clearly shows that the wave on 5 January is a wave of elevation. The observed negative perturbation at the surface is a manifestation of the fact that these waves are highly nonlinear with trapped regions of recirculating flow such as the one shown in Fig. 3. This calls into question earlier interpretations and classifications of solitary waves based strictly on surface-pressure signatures¹¹.

Our analyses of radiosonde data from Champaign and other

FIG. 1 Isochrons of the minimum surface pressure during the passage of the solitary waves on 5–6 January (solid lines) and 14 January (dashed lines) 1989, based on microbarogram traces from all National Weather Service surface observing stations in the central United States. Cities listed on the map correspond to microbarogram traces shown below the map. Isochrons are labelled in coordinated universal time (UT). Both solitary waves propagated from Missouri to Ohio, with measured phase speeds of 23 m s^{-1} (5 January) and 29 m s^{-1} (14 January) over central Illinois.

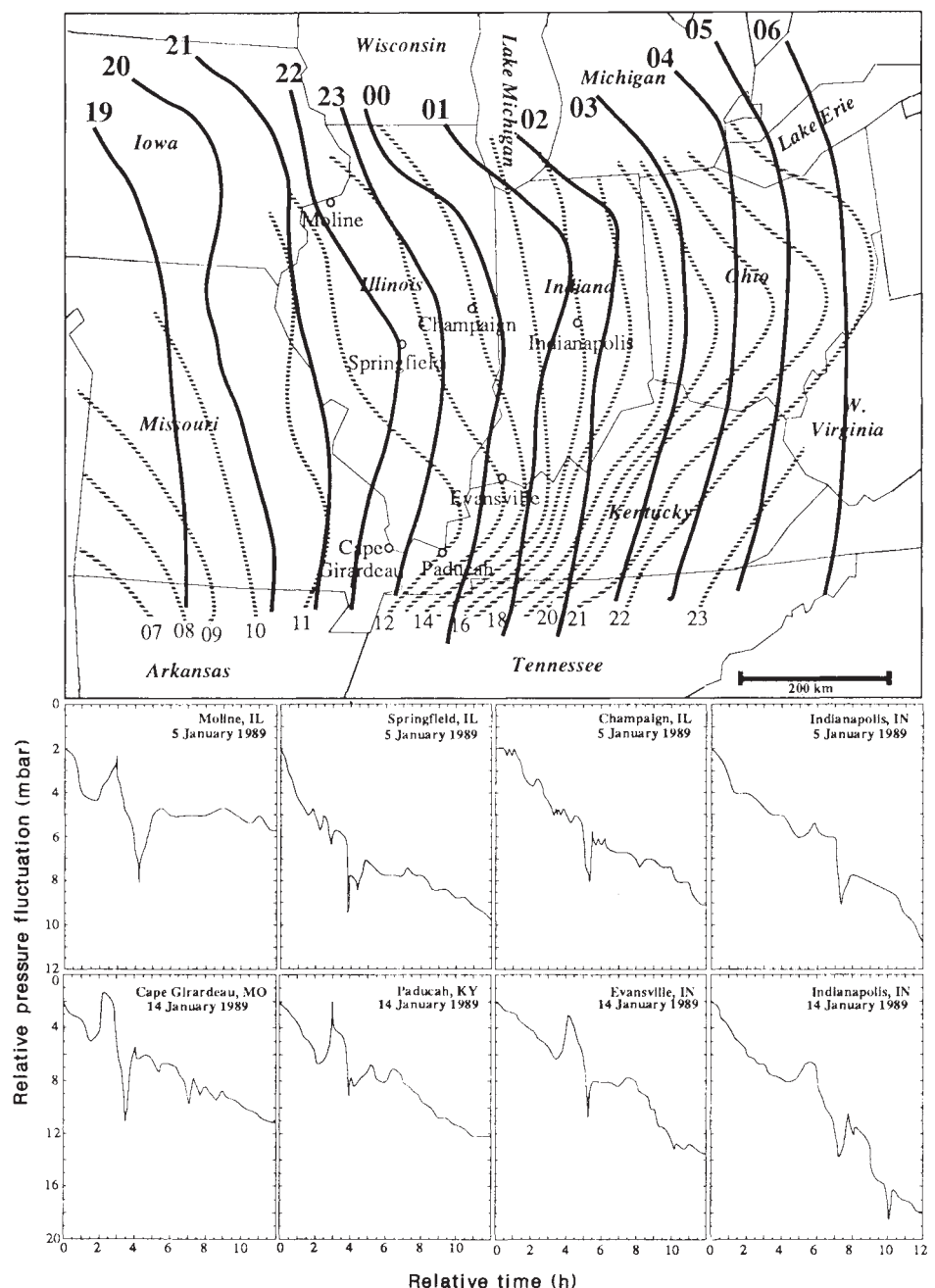
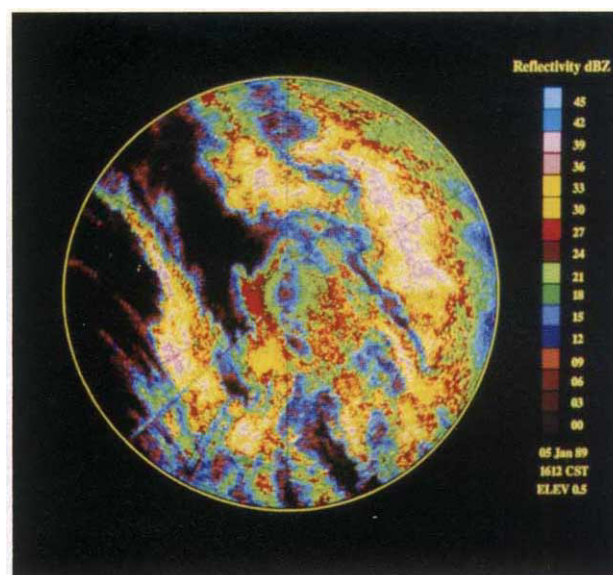


FIG. 2 Radar reflectivity factor²⁴ measured by the 10-cm-wavelength scanning Doppler radar on 5 January 1989 at 22:12 UT (16:12 Central standard time). The elevation angle of the scan was 0.5°. The diameter of the circle is 300 km with the radar located at the centre. The radar reflectivity factor Z is related to size of the particles intercepted by the radar beam by $Z = \sum_{vol} D^6$, where D is the diameter of a particle in a volume of air intercepted by the radar beam. Because of the large variation in Z , a logarithmic scale is used in which $dBZ = 10 \log_{10} Z$. In practice, the radar reflectivity factor is related to the intensity of the precipitation. The precipitation band, coincident with the elongated area of enhanced reflectivity on the left (west) side of the figure, is associated with the solitary wave.



locations at times before those shown in Fig. 3 indicate that there was a gradual lowering of the tropopause associated with the eastward propagation of the upper-level jet stream. During this earlier period, downward displacements of upper tropospheric air were also observed in the profiler data, indicative of inertia-gravity-wave activity. Inertia gravity waves associated with geostrophic adjustment alone, however, would not be

expected to evolve into a singular wave. Specific environmental conditions that promote wave trapping and amplification must be present. In the absence of a significant energy source, the prolonged existence of a solitary wave requires the presence of a waveguide from which very little energy is lost²². In the atmosphere, such a waveguide can be provided by a low-level temperature inversion and a critical level. A critical level, where

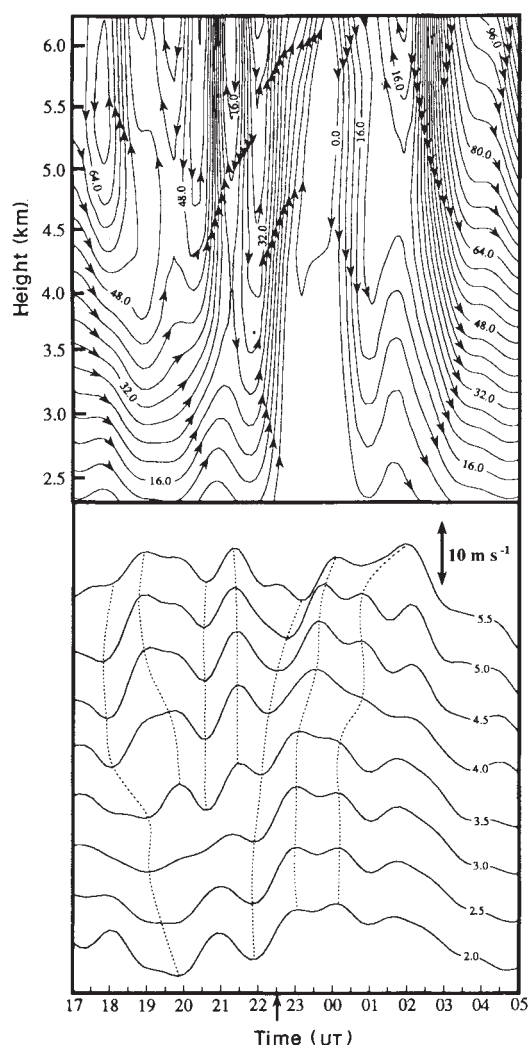


FIG. 3 Top, time-height cross-section of wave-normal relative streamlines Ψ ($10^{-3} \text{ m}^2 \text{ s}^{-1}$), with arrows indicating direction of flow, and bottom, time series of the u velocity component from 2.0 to 5.5 km for 5 January 1989. The bold arrow marks the time of minimum pressure during wave passage. The figure was constructed using wind-profiler observations from a location near Champaign. Because of the inherent limitations owing to the frequency used in this Doppler wind profiler, reliable observations are only available between 2 and 6 km. Dashed vertical lines depict lines of constant phase. Both Ψ and u were calculated from profiler data, with $\Psi = \int (u - c) dz$, u is the velocity component normal to the wavefront (m s^{-1}), and c is the phase speed of the wave (23 m s^{-1}), with a boundary condition of $\Psi = 0$ at 2 km. The data was filtered to eliminate perturbations with a period of less than 40 minutes. Just ahead of the solitary wave, strong lifting occurs, carrying the parcels from 2 km up to the critical level, near 6 km. On the back side of the wave, strong subsidence carries the parcels back down to 2 km. The $\Psi = 0$ streamline marks the boundary of a zone of recirculating fluid trapped in the wave between 4 and 6 km. The shift in the lines of constant phase is evidence of wave reflection off the critical level and down to the surface as the wave propagated eastward. Owing to the increase of wind speed with height, the lines of constant phase would tilt much more strongly when viewed with the abscissa chosen as distance rather than time.

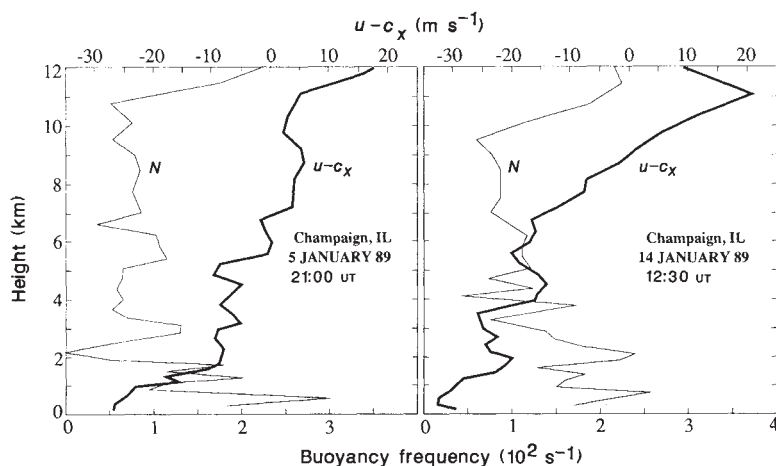


FIG. 4 Profiles of the moist Brunt-Vaisalla frequency N and $u - c_x$ taken from soundings at Champaign, ~2 h before solitary-wave passage on 5 and 14 January 1989. Here, $N = \sqrt{(g/\theta_e)(d\theta_e/dz)}$, where g is the gravitational force and θ_e is the equivalent potential temperature²⁵. Values of N , $u - c_x$ and Ri were calculated using radiosonde observations at Champaign near the time of wave passage. The slopes of N for both events generally decrease up to the tropopause level, whereas the $u - c_x$ profiles increase; the profile of 14 January shows more overall vertical shear. Three $u - c_x$ profiles within a 6-h period around the time of wave passage all had their zero crossings at the 6-km critical level on 5 January.

the wave-normal wind speed u exactly equals the phase speed of the wave c , acts as a reflector of wave energy if the Richardson number, $Ri = N^2/(\partial u/\partial z)^2$, is less than 0.25. Here N is the moist Brunt-Vaisalla frequency (see Fig. 4). In the duct, resistance to vertical displacements is minimized when the environment is neutrally stratified.

Figure 4 shows the profile of N and $u - c_x$ for both events ~2 h before the passage of the waves at Champaign. The inversion top on both days was located around 2 km, whereas the critical level was situated at about 6 and 9 km on 5 January and 14 January, respectively. These profiles lead us to believe that a well defined waveguide, which facilitated sustenance of the waves over a long distance, was present on both days. There also exists a great degree of similarity in the N and $u - c_x$ profiles. On both days, the wave-relative wind increases with height, whereas the moist Brunt-Vaisalla frequency decreases with height. Near the critical level, Ri was less than 0.25 on 5 January, a requirement for the over-reflection of vertically propagating gravity waves²³. The profile of Ri for 14 January differed only in that Ri was greater than 0.25 near the critical level. The observed phase speed of 23 m s^{-1} on 5 January matched the phase speed predicted by a linear model for ducted gravity waves²², within the uncertainties of the data used to perform the linear-model calculations.

Although these findings have shed some light on the charac-

teristics of large-amplitude solitary waves, more data is required to document the complete evolution of these events. Hopefully, a greater understanding of these occurrences will come when the planned network of wind profilers is deployed across the central United States in 1991. □

Geometric effects of deuteration on hydrogen-ordering phase transitions

M. I. McMahon*, R. J. Nelmes*, W. F. Kuhs†, R. Dorwarth†, R. O. Piltz* & Z. Tun*‡

* Physics Department, University of Edinburgh, Edinburgh EH9 3JZ, UK

† Institut für Kristallographie, Universität (TH) Karlsruhe, Postfach 6980, 7500 Karlsruhe 1, Germany

IT has long been known that the substitution of deuterium for hydrogen in hydrogen-bonded materials can lead to a change in the geometry of the hydrogen bonds—the so-called Ubbelohde effect¹. There can also be significant accompanying changes in the physical properties of the material. Among the most striking examples is the increase in the transition temperature, T_c , of hydrogen-ordering systems of the KH_2PO_4 type by >100 K on full deuteration. In all of these systems, T_c decreases under pressure, making it possible in principle to compare protonated and deuterated forms at the same T_c by applying pressure to the latter. Here we report the results of a high-pressure neutron diffraction study of the deuterated H(D)-ordering material PbDPO_4 . We find that much, and possibly all, of the increase in T_c on deuteration is attributable to the accompanying changes in the hydrogen-bond dimensions. This suggests that purely mass-dependent effects on the quantum-mechanical tunnelling between the equivalent H(D) potential minima—an explanation that has been favoured previously—have little or no direct influence on T_c . Substantial revisions to the theory of this type of ordering transition are apparently required.

KH_2PO_4 (KDP) undergoes an H-ordering phase transition to a ferroelectric phase on cooling through $T_c = 122 \text{ K}$ at atmospheric pressure. Above T_c , the H atoms are 50:50 disordered over two sites, δ apart, in short O—H—O bonds of length $2R$ linking the PO_4 tetrahedra. Below T_c , the H atoms order fully onto one of the sites, with accompanying displacements in the

Received 20 April; accepted 15 October 1990.

- Christie, D. R. & Muirhead, K. J. *Nature* **293**, 46–49 (1981).
- Lin, Y. H. & Goff, R. C. *J. Atmos. Sci.* **45**, 194–205 (1988).
- Fulton, R., Zrnic, D. S. & Doviak, R. J. *J. Atmos. Sci.* **47**, 319–337 (1990).
- Bosart, L. F. & Cussen, J. P. Jr *Mon. Wea. Rev.* **101**, 446–454 (1973).
- Bosart, L. F. & Sanders, F. J. *J. Atmos. Sci.* **43**, 924–939 (1986).
- Bosart, L. F. & Seimon, A. *Mon. Wea. Rev.* **116**, 1857–1886 (1988).
- DeMaria, M., Davis, J. M. & Wojcik, D. M. *Mon. Wea. Rev.* **117**, 826–842 (1989).
- Uccellini, L. W. & Koch, S. E. *Mon. Wea. Rev.* **115**, 721–729 (1987).
- Tepper, M. *J. Met.* **7**, 21–29 (1950).
- Smith, R. K. *Earth Sci. Rev.* **25**, 267–290 (1988).
- Stobie, J. G., Einaudi, F. & Uccellini, L. W. *J. Atmos. Sci.* **40**, 2804–2830 (1983).
- Pecnick, M. J. & Young, J. A. *J. Atmos. Sci.* **41**, 1850–1862 (1984).
- Koch, S. E. & Dorian, P. B. *Mon. Wea. Rev.* **116**, 2570–2592 (1988).
- Doviak, R. J. & Ge, R. J. *J. Atmos. Sci.* **41**, 2559–2573 (1984).
- Green, J. L. et al. *Geophys. Res. Lett.* **15**, 269–272 (1988).
- Wakimoto, R. M. *Mon. Wea. Rev.* **110**, 1060–1082 (1982).
- Rust, W. D. et al. *Bull. Am. met. Soc.* **71**, 173–180 (1990).
- Rottman, J. W. & Simpson, J. E. *Q. J. R. met. Soc.* **115**, 941–963 (1989).
- Cahn, A. Jr *J. Met.* **2**, 113–119 (1945).
- Christie, D. R., Muirhead, K. J. & Hales, A. L. *J. Atmos. Sci.* **35**, 805–825 (1978).
- Benjamin, T. B. *J. Fluid Mech.* **29**, 559–592 (1967).
- Lindzen, R. S. & Tung, K. K. *Mon. Wea. Rev.* **104**, 1602–1617 (1976).
- Rosenthal, A. J. & Lindzen, R. S. *J. Atmos. Sci.* **40**, 521–529 (1983).
- Battani, J. *Radar Observation of the Atmosphere* (University of Chicago Press, 1973).
- Iribarne, J. V. & Godson, W. L. *Atmospheric Thermodynamics* (Reidel, Dordrecht, 1973).

ACKNOWLEDGEMENTS. We thank the Aeronomy Laboratory of NOAA, particularly J. Green, for providing the profiler data, the Illinois State Water Survey for assistance in collecting the 10-cm radar data, and the National Center for Atmospheric Research for assistance with radiosonde data. This research was supported by the NSF.

‡ Present address: Physics Division, Chalk River Nuclear Laboratories, Chalk River, Ontario, Canada, K0J 1J0.

heavy-atom (K, P, O) structure. On deuteration to KD_2PO_4 (DKDP), T_c increases to 229 K—a ΔT_c of 107 K.

The quantum-tunnelling model was introduced² to account for the very large isotope effect on T_c , and this model has since been substantially developed to become the generally accepted basis for describing KDP-type H-ordering transitions³. The model attributes the ordering primarily to a transition in the interacting H-atom system itself, which localizes the H (or D) atoms into one of their two potential minima (inducing the heavy-atom displacements); quantum tunnelling opposes the localization, and hence T_c is lower in the H-form (KDP) because of the much higher tunnelling frequency of H. But the tunnelling model focuses on the mass change (from H to D) and takes no explicit account of the Ubbelohde effect. Accurate neutron diffraction studies⁴ have shown that this effect is far from negligible in KDP and that the increase in T_c is accompanied by significant increases in both δ and $2R$. Diffraction studies^{5,6} have also shown that both δ and $2R$ decrease with applied hydrostatic pressure, as does T_c (ref. 7), indicating that there may be some connection between the isotope effect on T_c and the accompanying structural changes.

A connection between T_c and structural dimensions is not a new idea. A possible dependence of T_c on δ , of the form $T_c \propto \delta^2$, was suggested some time ago by Silsbee *et al.*⁸ on the basis of theoretical considerations. More recently, Ichikawa and colleagues (ref. 9 and earlier work cited therein) have used crystal-structure results on KDP and several of its isomorphs to suggest an empirical proportionality between T_c and $2R$, implying $\Delta T_c = 0$ K at constant $2R$. Samara and Semmingsen¹⁰ reached the same conclusion for the system $\text{H}_2\text{C}_4\text{O}_4$ – $\text{D}_2\text{C}_4\text{O}_4$ (which has a different structure from KDP) on the basis of an estimated compressibility for $2R$, but also argued that the same should not be true for KDP where (unlike $\text{H}_2\text{C}_4\text{O}_4$) tunnelling was expected to have a major role in deuteration effects. The values and estimates of $2R$ used in these papers have since been shown to be insufficiently accurate to support fully the conclusions drawn^{11,12}, but the notion of some dependence of T_c on $2R$ remains (δ was not considered).

To investigate more fully the possible relationship between T_c and structural dimensions in the KDP–DKDP system, it would be necessary to determine the crystal structures of KDP and DKDP with them both at the same value of T_c . But a pressure of ~40 kbar is needed to reduce T_c in DKDP to ~122 K (ref. 7), and this is as yet unobtainable using single-crystal neutron diffraction techniques. (No other technique can be used to determine details of a H(D) distribution.) Hence, our conclusions so far have depended on extrapolations—in this case from measurements made up to 17 kbar^{5,6}. But we can now report the results obtained from another system with the KDP-type H-ordering transition, PbHPO_4 – PbDPO_4 (LHP–LDP), in which the protonated and deuterated materials can be studied at the same T_c .

LHP undergoes an H-ordering transition at 310 K, substantially higher than in KDP, and possesses a very different crystal structure: in KDP, the H-bonds link PO_4 groups into a three-dimensional network, whereas in LHP the PO_4 groups are linked only along one direction, forming HPO_4 chains¹³ (Fig. 1). Des-

TABLE 1 Structural changes in PbDPO_4 and PbHPO_4 with P and T

	PbDPO_4	PbHPO_4	PbDPO_4	PbHPO_4
T (K)	330 (2)	316 (1)	457 (1)	230 (1)
P (kbar)	13.8 (5)	AP*	AP*	6.9 (8)
T_c (K)	308 (10)	310 (1)	452 (1)	221 (4)
a (Å)	4.6430 (4)	4.6829 (2)	4.691 (1)	4.656 (1)
b (Å)	6.5669 (7)	6.6447 (2)	6.673 (2)	6.581 (4)
c (Å)	5.7434 (5)	5.7798 (2)	5.800 (1)	5.749 (1)
β (deg)	96.748 (3)	97.150 (3)	97.21 (1)	96.85 (3)
δ (Å)	0.402 (7)	0.391 (5)	0.476 (5)	—
$2R$ (Å)	2.460 (2)	2.470 (1)	2.496 (1)	2.449 (2)
θ (deg)	28.5 (3)	25.9 (1)	26.1 (5)	27.1 (1)

Estimated standard deviations on the last quoted place are given in parentheses. The crystal structure remains slightly acentric above T_c (ref. 22), and the H(D)-bond dimensions and θ in columns 1, 2 and 3 were calculated from the fractional coordinates obtained from refinements of PbHPO_4 and PbDPO_4 in the acentric spacegroup Pc . The data for PbHPO_4 at 6.9 kbar are of insufficient quality to allow refinement of the small acentricity and were therefore refined in the corresponding centrosymmetric spacegroup $P2_1/c$; the data are also of insufficient quality to allow an accurate determination of δ . The relatively large errors on the unit-cell dimensions (a, b, c, β) and θ in the atmospheric-pressure study of PbDPO_4 are due to uncertainties in the neutron wavelength used. θ is the angle $P'-P-O_1$ projected onto the ac plane and characterizes the orientation of the PO_4 group around the b axis (see Fig. 1).

* AP, atmospheric pressure.

pite these differences, the isotope effect on T_c is similar to that in KDP, with T_c rising to 452 K in LDP¹³—a ΔT_c of 142 K. Hydrostatic pressure decreases T_c in LHP and LDP much more strongly than in KDP: we have measured dT_c/dP to be about -11 K kbar⁻¹ in LDP, similar to the value for LHP¹⁴. Thus only a relatively modest pressure of 14 kbar is required to reduce T_c in LDP to 310 K. This pressure is within the range of a newly developed sapphire-anvil pressure cell¹⁵ specifically designed for accurate single-crystal neutron diffraction studies of low-symmetry structures.

We collected the neutron diffraction data at the Institut Laue-Langevin (Grenoble) using high-resolution single-crystal techniques. The single-crystal sample of LDP was maintained at 13.8 kbar and 330 K using resistive heating in the pressure cell. T_c was determined to be 308 ± 10 K, the same as that for LHP (310 ± 1 K) within the uncertainty. (The error of 10 K on T_c reflects the relative error in the determination of sample pressure, affecting the estimate of T_c through dT_c/dP .) We collected the data to $\sin \theta/\lambda = 1.1 \text{ \AA}^{-1}$ at a calibrated wavelength of $0.7071(1) \text{ \AA}$, and corrected for pressure-cell and sample absorption and for anisotropic extinction. All corrections and subsequent least-squares refinements were made with the Promethues suite of crystallographic programs¹⁶. Table 1 presents the measured unit-cell dimensions, the values of δ and $2R$ derived from the refined atomic coordinates, and an angle θ characterizing the orientation of the PO_4 groups relative to the HPO_4 chain direction (see the caption). The table also includes results obtained in our atmospheric-pressure studies of LHP at 316 K ($T_c + 6$ K) and LDP at 457 K ($T_c + 5$ K), and in a preliminary

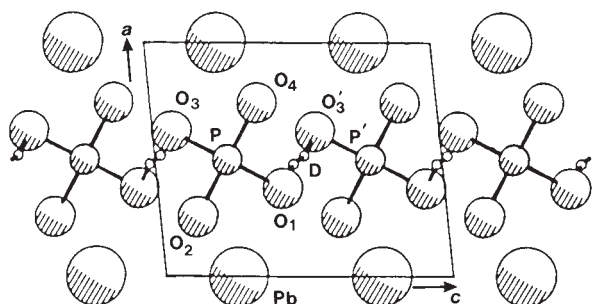
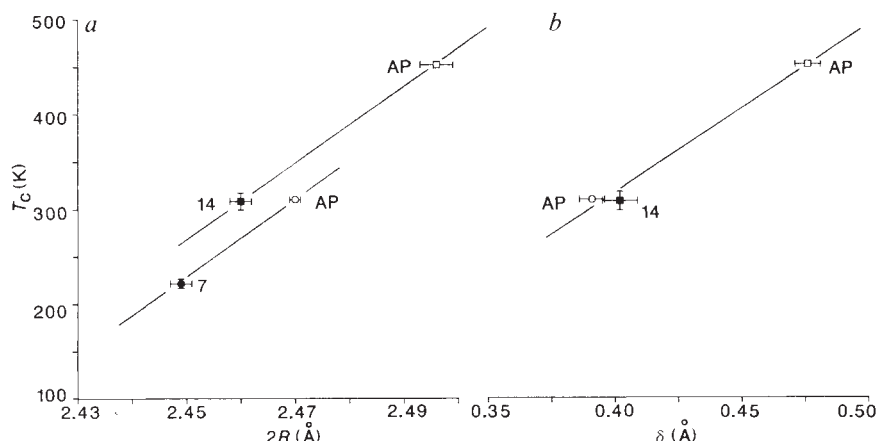


FIG. 1 The crystal structure of PbDPO_4 above T_c , viewed down the b axis. The two formula units in the primitive unit cell are related by a c glide plane perpendicular to the b axis. With the origin in the glide plane, the approximate b -axis fractional coordinates of the labelled atoms are 0.21 (Pb), 0.79 (P), 0.65 (O_1, O_3), 0.92 (O_2, O_4), 0.34 (O'_1) and 0.21 (P'). The D atoms are disordered over two sites, as shown, in the O—D—O bonds that link the PO_4 groups along the c axis.

FIG. 2 *a*, The variation of T_c with the O—H(D)—O bond length $2R$ and *b*, the variation of T_c with the H(D) site separation δ under pressure in PbHPO_4 and PbDPO_4 . The PbHPO_4 points are represented by circles, and the PbDPO_4 points by squares; the high-pressure points are distinguished by filled-in symbols. The pressure is given against each point as AP (atmospheric pressure) or the number of kilobars applied. Where no error bar is given, the error is less than the size of the symbol used to plot the point. The lines are guides to the eye only.



study of LHP just above T_c at 7 kbar—all obtained by single-crystal neutron diffraction techniques¹⁷.

The values of $2R$ are plotted against T_c in Fig. 2*a*. If LHP and LDP are compared at the same $2R$ (for example, at 2.470 Å), ΔT_c is greatly reduced from its atmospheric-pressure value of 142 K, but is still ~40 K. However, if the comparison is made at the same H(D)—H(D) separation δ (Fig. 2*b*), ΔT_c is reduced to zero within 10 K: LHP and LDP seem to have the same T_c at the same δ . (The value of ~40 K for ΔT_c at constant $2R$ is a correction to the 108 K given in a preliminary account¹⁷.)

These results clearly support the conclusions made by extrapolation in the KDP-DKDP system¹¹. In that case also, ΔT_c with $2R$ held constant at its atmospheric-pressure value in KDP was estimated as 40 K (certainly not zero), whereas at constant δ the value of ΔT_c was estimated as only 15 K—possibly zero within the uncertainties. Although the KDP-DKDP results were less well determined, it thus does seem that KDP-DKDP and LHP-LDP—two very different H-ordering systems—share a remarkably similar dependence of ΔT_c on both $2R$ and δ . This contrasts with the assumptions underlying the earlier study, referred to above, of deuteration effects in the $\text{H}_2\text{C}_4\text{O}_4$ – $\text{D}_2\text{C}_4\text{O}_4$ system¹⁰.

Deuteration at constant $2R$ is expected to correspond most closely to replacing H by D in the same potential¹⁸. The ΔT_c this gives (~40 K) is only a third of the change the tunnelling model was originally devised to explain. This could be understood in terms of tunnelling having two effects¹⁹: a direct effect (here 40 K) as predicted by the tunnelling model but of considerably lesser magnitude, and an indirect effect through the influence of tunnelling on the H-bond dimensions. But it is also possible that δ is the principal determinant of T_c (ref. 20); indeed the balance of structural evidence favours this interpretation, although not yet decisively^{17,20}. In this case there is no significant direct tunnelling effect—the large increase in T_c at atmospheric pressure is entirely, or almost entirely, attributable to the accompanying increase in δ ; it is this increase in δ that emerges as the primary effect of deuteration. (If T_c does depend directly

on δ , the change in T_c on deuteration at constant $2R$ is readily understood: at constant $2R$, the double-minimum H(D) potential remains unaltered as H is replaced by D, but the D energy levels are lower in the potential minima¹⁸; owing to the asymmetry of the minima, δ is then larger for D than for H at the same $2R$, giving rise to the increase in T_c .) δ could determine T_c along the lines suggested by Silsbee *et al.*⁸, although our diffraction results²⁰ on KDP and $\text{H}_2\text{C}_4\text{O}_4$ indicate that T_c varies linearly with δ rather than as δ^2 , and δ remains non-zero as T_c decreases to 0 K (unpublished data). In any case, it may not be fruitful to search for an understanding within the existing framework in view of the possibility that transitions of this kind are 'driven' by interactions in the heavy-atom system rather than in the H(D) system²¹. If this is the case, the effect of δ on the geometry of the H-bonded heavy-atom units (PO_4 , for example) would be the important factor. We need a revised theory of this type of transition to determine whether, and for what reason, $2R$ or δ is the more significant structural dimension, and to account for the observed common values of ΔT_c . Some interesting new ideas are now emerging^{21,23}.

There is another important point to consider in interpreting deuteration effects. When δ is the same (Table 1, columns 1 and 2), $2R$, θ and the unit-cell dimensions are significantly different. Interpolating between columns 1 and 3 to the point where $2R$ is the same, shows that then δ , θ and the cell dimensions are different. The same is true of the KDP-DKDP system. Thus, it is not possible to determine the effect of simply replacing H by D in a constant atomic arrangement. If one dimension (like δ) is kept constant, there are always other geometrical changes (including those in the heavy-atom system), and these may mask or enhance the effect of deuteration alone.

Our evidence that δ may be the principal determinant of T_c invites a revised theory of H-ordering transitions that takes explicit account of geometric effects. The fact that it is not generally possible to deuterate in an otherwise constant structure is a point of wider significance for the use and interpretation of deuteration. □

Received 9 July; accepted 1 October 1990.

- Robertson, J. M. & Ubbelohde, A. R. *Proc. R. Soc. A* **170**, 222–240 (1939).
- Blinc, R. J. *Phys. Chem. Solids* **13**, 204–211 (1960).
- Tokunaga, M. & Matsubara, T. *Ferroelectrics* **72**, 175–191 (1987).
- Tun, Z., Nemes, R. J., Kuhs, W. F. & Stansfield, R. F. D. *J. Phys. C* **21**, 245–258 (1988).
- Tibbals, J. E., Nemes, R. J. & McIntyre, G. J. *J. Phys. C* **15**, 37–58 (1982).
- Tibbals, J. E. & Nemes, R. J. *J. Phys. C* **15**, 849–853 (1982).
- Samara, G. A. *Ferroelectrics* **71**, 161–182 (1987).
- Silsbee, H. B., Uehling, E. A. & Schmidt, V. H. *Phys. Rev. A* **133**, 165–170 (1964).
- Ichikawa, M., Motida, K. & Yamada, N. *Phys. Rev. B* **36**, 874–876 (1987).
- Samara, G. A. & Semmingsen, D. *J. chem. Phys.* **71**, 1401–1407 (1979).
- Nemes, R. J. *J. Phys. C* **21**, L881–L886 (1988).
- Katrusiak, A. & Nemes, R. J. *J. Phys. C* **19**, L765–L772 (1986).
- Negran, T. J. *et al. Ferroelectrics* **6**, 179–182 (1974).

- Mylov, V. P., Shirokov, A. M., Shuvalov, L. A., Kharitonov, V. N. & Brzhnezina, B. *Soviet Phys. Crystallogr.* **24**, 738–739 (1979).
- Kuhs, W. F., Ahsbals, H., Londono, D. & Finney, J. L. *Physica B* **156/157**, 684–687 (1989).
- Zucker, U. H., Perenthaler, E., Kuhs, W. F., Bachmann, R. & Schulz, H. *J. appl. Cryst.* **16**, 358 (1983).
- Piltz, R. O., McMahon, M. I. & Nemes, R. J. *Ferroelectrics* **108**, 271–276 (1990).
- Lawrence, M. C. & Robertson, G. N. *J. Phys. C* **13**, 1053–1059 (1980).
- Matsubara, T. & Matsushita, E. *Prog. theor. Phys.* **71**, 209–11 (1984).
- McMahon, M. I., Piltz, R. O. & Nemes, R. J. *Ferroelectrics* **108**, 277–282 (1990).
- Kojyo, N. & Onodera, Y. *J. phys. Soc. Japan* **57**, 4391–4402 (1988).
- Lockwood, D. J., Ohno, N., Nemes, R. J. & Arend, H. *J. Phys. C* **18**, L559–L565 (1985).
- Motida, K. & Ichikawa, M. *J. phys. Soc. Japan* **58**, 2810–2816 (1989).

ACKNOWLEDGEMENTS. We thank H. Ahsbals for assistance in preparing the high-pressure cell, H. Arend and D. J. Lockwood for supplying the sample crystals, and SERC and the Bundesministerium für Forschung und Technologie for support.

Deep-water renewal by turbidity currents in the Sulu Sea

Detlef Quadfasel*, Hermann Kudrass†
& Andrea Frische*

* Institut für Meereskunde, Universität Hamburg,

2000 Hamburg 54, Germany

† Bundesanstalt für Geowissenschaften und Rohstoffe,

3000 Hannover 51, Germany

THE ventilation of the intermediate and deep layers of the ocean is governed by convection: dense water formed at the sea surface through air-sea interaction processes (mainly at high latitudes) sinks to great depths. Convection determines the heat, salt and dissolved-gas budgets of the ocean's interior and therefore exerts a significant influence on the Earth's climate. Convective processes can also be initiated by the flow of dense water over submarine sills separating ocean basins with water masses of different buoyancy. Here we report the observation of relatively light deep water in the overflow region near the sill in the northern part of the 5,000-m-deep Sulu Sea. Because of its buoyancy, this water could not have reached the sea floor of its own accord. Turbidity currents, induced by changes in water density that result from sediment influx, occur in this region at intervals of several decades, as indicated by the frequent deposition of graded sequences of silty mud on the abyssal plain. We suggest that such turbidity currents, combined with the effects of classical hydrological plume convection, are capable of bringing light water to large depths. In tropical regions, where normal oceanic convection is relatively weak, this mechanism may contribute significantly to the ventilation of deep basins.

The Sulu Sea in the eastern Indonesian archipelago is a small, 5,000-m-deep basin surrounded by shallow shelves and islands. Its deep water below a depth of ~2,000 m is remarkably homogeneous in temperature and salinity. Geochemical tracer data indicate renewal times of the deep water of 50–100 years¹. The deep water is replenished by dense thermocline water from the South China Sea, spilling over the 420-m-deep sill from the Mindoro Strait² into a broad channel running parallel to the western side of Panay Island (Fig. 1). Other channels providing a connection between the Sulu Sea and the surrounding seas are shallower than 200 m and thus cannot be pathways for inflowing waters capable of sinking to large depths². In the southern part of the Mindoro Strait the deep channel steepens forming a canyon, which opens to the abyssal plain at a water depth of ~4,000 m. The plain slopes southward down to the maximum depth of 5,000 m in the central part of the Sulu Sea.

During cruise 58 of RV *Sonne*^{3,4} in August 1988, vertical profiles of temperature, salinity, dissolved oxygen content and nutrients were obtained along a section from the Manila trench in the South China Sea to the southern Sulu Sea (Fig. 1). In addition, 34 piston cores with an average length of 8 m were collected during this cruise and one in 1987.

The vertical distribution of potential temperature and salinity along the section (Fig. 2) shows the spreading of water masses from the South China Sea into the Sulu Sea. The core of Subtropical Lower Water from the Pacific is marked by a distinct salinity maximum and lies below the warm and fresh surface layer around 200 m depth, well above the sill depth in the Mindoro Strait. Its spreading is thus not hampered by topography. Northern Pacific Intermediate Water, characterized by a salinity (and oxygen) minimum at temperatures between 7.5 and 9.5 °C, is clearly visible in the South China Sea at depths between 400 and 800 m. This water is the primary source for the intermediate and deep waters of the Sulu Sea². A low-salinity core ($S < 34.46$) originating from this minimum can be traced across the Sulu Sea in the depth layer 600–1,400 m. Temperatures and salinities here are higher than in the corresponding layer

TABLE 1 Mean hydrographic properties in the Sulu Sea

Depth (m)		Potential temperature (°C)		Salinity		Density		Dissolved oxygen (mg l ⁻¹)	
		Θ	$\delta\Theta$	S	δS	σ_2	$\delta\sigma_2$	O ₂	δO_2
2,000	N	9.887		34.465		35.377		1.59	
	S	9.883	0.004	34.465	0.000	35.378	-0.001	1.50	0.09
3,000	N	9.882		34.474		35.385		1.32	
	S	9.876	0.006	34.478	-0.004	35.387	-0.002	1.21	0.11
4,000	N	9.881		34.474		35.385		1.30	
	S	9.875	0.006	34.476	-0.002	35.387	-0.002	1.30	0.000

N, northern stations 35, 40 and 73. S, southern stations 51, 77, 83 and 107.

in the South China Sea, owing to tidally induced mixing in the Mindoro Strait and entrainment^{5,6} of ambient water during its descent down the northern Panay Canyon. The high oxygen concentrations in this layer ($>2.5 \text{ mg l}^{-1}$)—high compared to data collected in 1976¹ when maximum values in this layer were $<2.2 \text{ mg l}^{-1}$ —indicate that this intrusion occurred relatively recently.

In the deep-water regime below 2,000 m, the vertical and horizontal gradients of the hydrographic properties are small (Fig. 2). There is a tendency for the deep water in the northern Sulu Sea to be slightly fresher and warmer and also to have a higher oxygen content compared to the deep water in the south (Table 1). The observed differences are marginal with respect to the accuracy of the measurements, but consistent at all depth

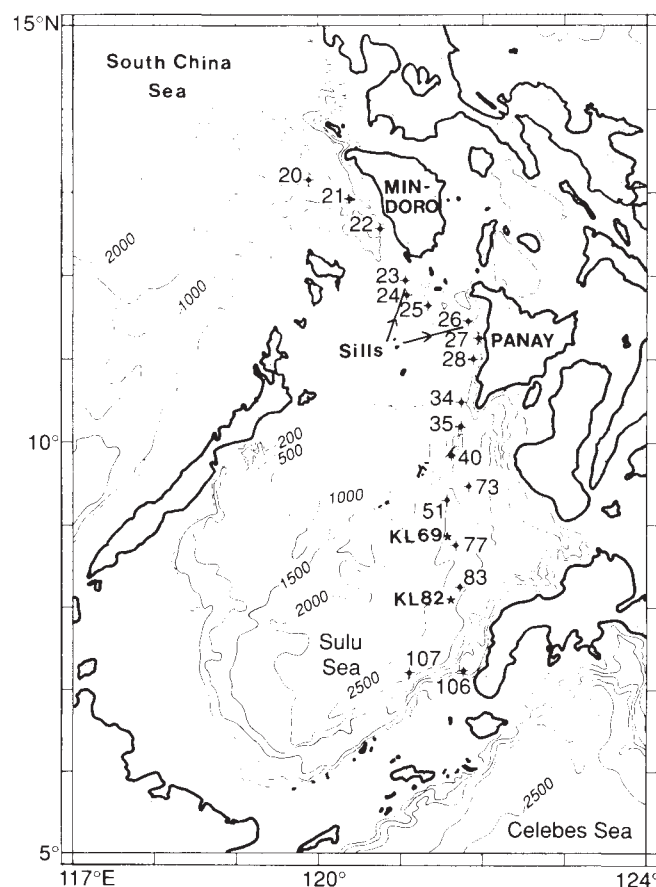


FIG. 1 Map of the Sulu Sea with the location of the hydrographic stations (28 August–17 September 1988) and location of the two sediment coring sites 69KL and 82KL. Depth contours in fathoms (1.82 m).

levels and for all properties and indicate an intrusion of Northern Pacific Intermediate Water in the northern Sulu Sea. Plume convection^{5,6} down the slope in the northern Panay Canyon is a possible mechanism.

In the purely oceanic case an initially very dense water mass sinks down a slope. On its way, the plume entrains ambient water and thus loses more and more of its original characteristics. The sinking process stops once the plume has attained the same density as the ambient water. Our observations are not consistent with this, however, as the water mass near the slope is less dense than the deep ambient water in the southern basin.

An explanation for this is that the overflow water gets loaded with suspended sediments, increasing its density, and plunges down the slope in the form of a turbidity current⁷. A layer of <0.1 cm of sediment suspended in a 100-m water column would make the density of the plume greater than that of the deep water in the central basin. Once the sediment has settled, the light water will rise as buoyant plumes and mix with the overlying denser water.

The laminated sediments in the abyssal plain of the Sulu Sea provide convincing evidence for frequent turbid-plume convective events^{8,9}. One sediment core in particular (SO49-82KL, taken at the western flank of the nearby 4,980-m-deep abyssal

plain) is suitable for an estimation of the frequency of turbidity currents, as it contains a 6-m-long Holocene record of turbidity currents and of hemipelagic sedimentation. Since 10,000 yr BP (before present), 142 turbidity currents have deposited graded, finely laminated silt or mud layers (Fig. 3). During this period the median thickness of individual turbidites decreased from 36 mm in the early Holocene to 19 mm in the late Holocene. The average interval between these events also changed from 51 yr between 10,000 and 7,120 yr BP to 83 yr between 7,210 yr and today, probably reflecting the changes of the sea level. In another core (SO58-69KL, water depth 4,696 m, location shown in Fig. 1), taken ~200 m above the abyssal plain, 150 turbidites have been counted for the past 10,000 years occurring at an average frequency of 67 yr. Presumably not all turbidity currents passing to the abyssal plain leave behind a sedimentary record on the slope surrounding the plain and the frequency of these events may therefore be higher than the measured one. The thickness of individual turbidites increases considerably from the flanks to the abyssal plain. Unfortunately these cores could not be dated as they contain numerous erosional hiatuses.

What mechanism triggers the turbidity-flow events and the replenishment of the deep water? Earthquakes combined with slumping may cause turbidity flows. But the frequency of tur-

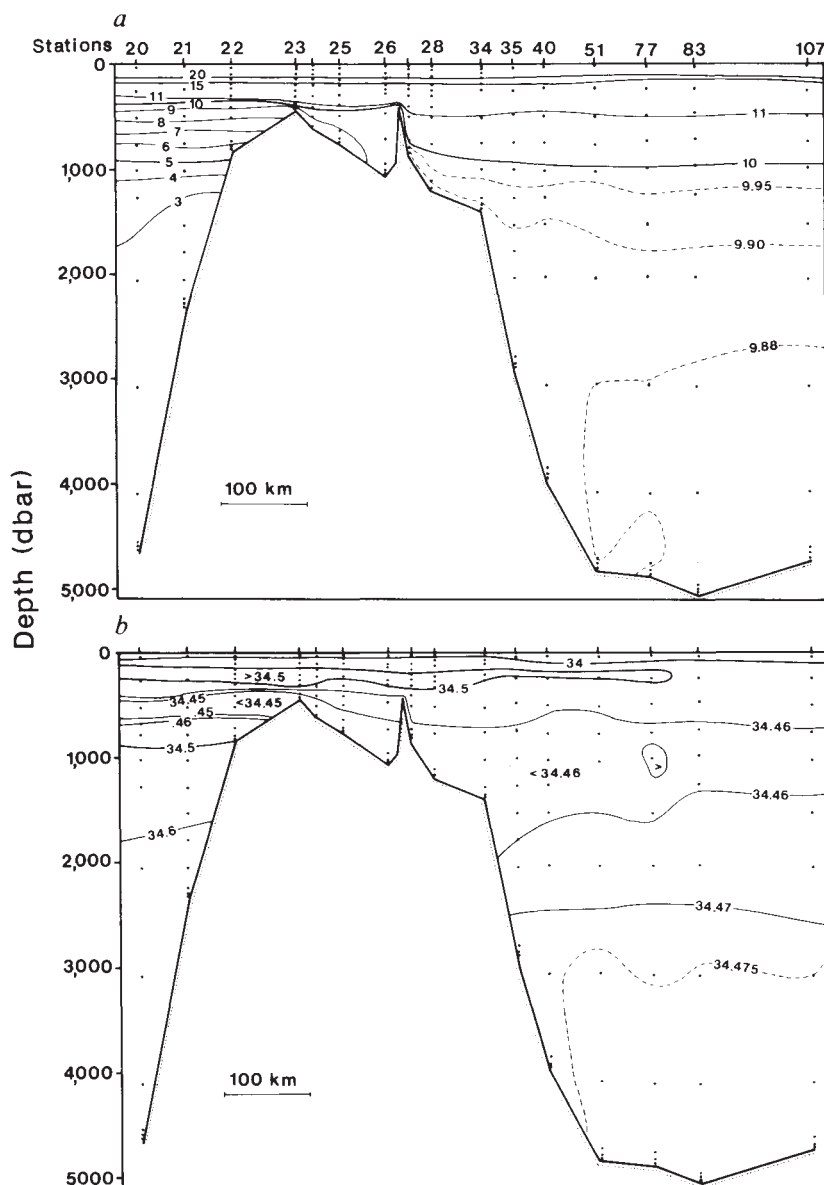
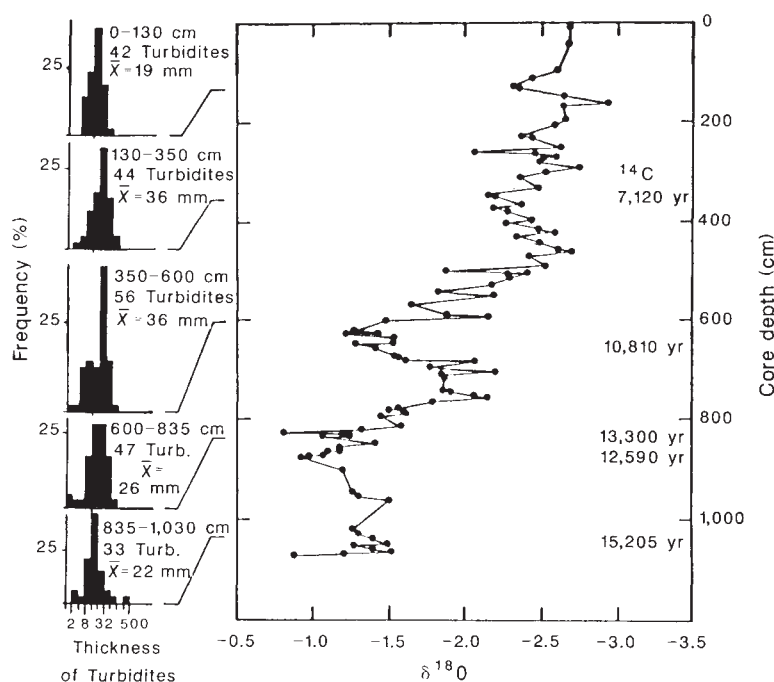


FIG. 2 Vertical distribution of *a*, potential temperature and *b*, salinity along the hydrographic section. See Fig. 1 for location.

FIG. 3 Oxygen isotope record of *Globigerinoides ruber* of core SO49-82KL (water depth 4,911 m, see Fig. 1 for location). The ^{14}C dates of planktonic foraminifera ($>125\ \mu\text{m}$) have been corrected for sea water by subtracting 400 years from the measured radiocarbon ages¹¹. The thickness of turbidites was measured on radiograph slides. $\bar{x}$ indicates the median thickness of turbidites.



bidites in a sub-basin lying west of the studied abyssal plain and adjacent to the central ridge of the Sulu Sea (Fig. 1) is more than an order of magnitude lower. As earthquakes have a relatively large area of influence, they are probably not causing the frequent turbidite signal found in the abyssal plain. The South China and Sulu Seas are prominent pathways of tropical cyclones, however, and in the wake of cyclones strong uplifting of the thermocline has been observed¹⁰. When denser water is brought up above the sill depth of the Mindoro Strait, this water will overflow and erode sediments thereby enforcing the downward motion of the plume. Thus, two conditions have to be met for deep plume convection to occur: there has to be enough sediment available in the strait and there has to be an initial overflow strong enough to erode the sediments. The timescales associated with these processes are different. Strong cyclones pass the Mindoro Strait about every five to ten years but the sedimentation owing to river input is a continuous process. Typical sedimentation rates off the mouths of tropical rivers are $0.5\text{--}1\ \text{cm yr}^{-1}$. Thus with an average interval of 50 years between convective events, as estimated from the core samples, at least 25 cm of sediment has to be accumulated to provide the critical mass for this convection to occur.

The convection caused by turbidity currents will leave the waters of the deep Sulu basin less dense and thus lower the threshold for more normal (oceanic) density-driven plume convection. They essentially lower the residence time of the deep basin waters by enabling a more rapid ventilation. The relative contribution of turbidity currents to the total ventilation of the deep waters in the ocean cannot be estimated from our single synoptic hydrographic section or from the sediment records. Measurements during an actual convective event would be necessary to quantify the amount of water brought down in the plumes. □

10. Sandford, T. B. *et al. J. phys. Oceanogr.* **17**, 2065–2083 (1987).

11. Broecker, W. S. *et al. Nature* **333**, 156–158 (1988).

ACKNOWLEDGEMENTS. We thank H. Erlenkeuser, M. Geyk and U. Brockmann for the analysis of oxygen isotopes, ^{14}C dates and nutrients. This work was funded by Bundesminister für Forschung und Technologie and by Deutsche Forschungsgemeinschaft.

Dissolved titanium in the open ocean

Kristin J. Orians*[‡], Edward A. Boyle*
& Kenneth W. Bruland[†]

* Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[†] Institute of Marine Sciences, University of California, Santa Cruz, California 95064, USA

VERY little is known about the marine geochemistry of dissolved titanium, mainly because of the difficulties involved in the analysis of this highly refractory element at the extremely low levels present in sea water. Recent advances in analytical instrumentation using inductively coupled plasma mass spectrometry have lowered trace-element detection limits (particularly for refractory metals) to levels that occur naturally^{1,2}. Titanium is relatively abundant in the Earth's crust (0.57% by weight³), yet its concentration in sea water is extremely low. Here we present measurements of titanium profiles in the open ocean. Dissolved titanium is found to be a reactive short-residence-time element, with a unique and highly non-uniform spatial distribution. Dissolved titanium is depleted in surface waters and enriched in deep waters, with a range of more than two orders of magnitude, and there are several indications that it is scavenged (removed by biotic or abiotic processes). This unusual oceanic distribution makes titanium potentially useful as a new tracer of chemical transport processes in deep waters.

Thermodynamic models of trace-element speciation^{4,5} predict that the neutral species, $\text{TiO}(\text{OH})_2$, should be the dominant form of dissolved titanium in sea water. This species is expected to be particle reactive^{6,7}, resulting in a short oceanic residence

Received 19 February; accepted 5 October 1990.

1. Broecker, W. S., Patzert, W. C., Toggweiler, J. R. & Stuiver, M. *J. geophys. Res.* **91**, 14345–14354 (1986).
2. Wyrtki, K. Naga Rep. Vol. 2 (Scripps Institution of Oceanography, La Jolla, 1961).
3. Vollbrecht, R. & Kudrass, H. R. *Init. Rep. ODP* **124/124E**, 105–111 (1990).
4. Frische, A. & Quadfasel, D. *Init. Rep. ODP* **124/124E**, 101–104 (1990).
5. Smith, P. C. *Deep Sea Res.* **22**, 853–874 (1975).
6. Quadfasel, D., Rudels, B. & Kurz, K. *Deep Sea Res.* **35**, 1143–1150 (1988).
7. Heezen, B. C. in *The Sea Vol. 3* (ed. Hill, M. N.) 742–775 (Wiley, New York, 1963).
8. Exon, N. F. *et al. Mar. Geol.* **39**, 165–195 (1981).
9. Wetzfel, A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **42**, 285–304 (1983).

[‡] Present address: Department of Oceanography, University of British Columbia, Vancouver, British Columbia, Canada, V6T 1W5.

time. Titanium is therefore potentially analogous to aluminium, thorium, iron and other hydroxide-dominated elements, a group of elements recently receiving increased attention⁸⁻¹². Iron is of particular interest owing to its potential importance as a bio-limiting nutrient^{11,12}. Elucidation of the biogeochemistry of titanium in the oceans will add to our understanding of this series of short-residence-time elements. The initial examination of dissolved titanium in the open ocean presented here allows a first-order interpretation of the biogeochemical processes that control its distribution in the marine environment.

Samples from the North Pacific were collected in the spring and summer of 1987 (as part of the VERTEX program) from a station at the edge of the California current (33° N, 139° W), and a station in the Sub-Arctic current (50° N, 145° W, Ocean Station 'P' or PAPA). We used sampling techniques^{13,14} designed to minimize trace-element contamination. Seawater samples were filtered, acidified, stored and processed as for dissolved gallium¹⁵. A subsample of this eluant was later diluted, a ⁴⁹Ti spike was added, and this was processed through the 8-hydroxyquinoline resin¹⁶ to separate potential interferents (sulphate and phosphate) from titanium, using a potassium acid phthalate rinse, before elution with 2.4M HNO₃. Samples from the North Atlantic were collected in February 1989, off Bermuda (32° N, 64° W, near Station 'S'), using acid-cleaned, modified Niskin bottles (epoxy-coated internal springs, silicon O-rings) and standard hydro-wire. Seawater samples were filtered, acidified, spiked with ⁴⁹Ti, adjusted to pH 3.5–4.0, pumped through the 8-hydroxyquinoline resin, rinsed with H₂O and potassium hydrogen phthalate, and eluted with 2.4M HNO₃. Titanium was detected by inductively coupled plasma mass spectrometry (ICP-MS) using ⁴⁹Ti isotope dilution analysis (details of the technique will be published elsewhere). We achieved a 400–500-fold concentration, with separation from Ca²⁺, SO₄²⁻, PO₄³⁻ and other major ions in sea water, resulting in a detection limit of 4 pM titanium for a 5-l seawater sample. This compares with published detection limits of 30–60 pM (ref. 17), which are an order of magnitude too high for the detection of the extremely low levels found in the surface waters of the North Pacific. The sampling and analytical precision, based on percent difference for three samples collected and analysed in duplicate (at levels of 90–200 pM), is ±10%.

The distribution of dissolved titanium in the North Pacific reflects a complex combination of biogeochemical controls. Surface levels are low, 4–8 pM, and increase nearly linearly with depth to a maximum near bottom of 200–300 pM (Fig. 1a). The vertical distribution of dissolved titanium in the Sub-Arctic North Pacific is markedly different than that of other scavenged-type, hydrolysis-dominated elements, such as aluminium, which typically have surface maxima from aeolian sources and low concentrations at depth owing to rapid removal by scavenging onto particles (Fig. 1b). The vertical distribution of titanium is also in marked contrast to nutrient-type elements, such as silica or phosphorus (Fig. 1d). The shape of the dissolved titanium profile at this site is most similar to that of dissolved copper (Fig. 1c). The dynamic range for titanium, however, is greater than two orders of magnitude whereas that of copper is only a factor of three or four. Copper is controlled by a complex combination of biological cycling and particle reactivity^{14,18,19}. The resemblance of the two profiles could suggest control by similar oceanographic processes for copper and titanium.

Dissolved titanium concentrations above 2,000 m at the station on the edge of the California current, at 33° N, are not significantly different from the levels observed further north, in the Sub-Arctic current, at 50° N (Fig. 2). This is surprising in light of the variation observed in other trace-metal levels with latitude in this region. The concentrations of aluminium and gallium both decrease by a factor of two to three, in the higher latitudes, primarily from increased scavenging in the highly productive Sub-Arctic current¹⁵. Nutrient-type elements and dissolved copper, on the other hand, have increased concentra-

TABLE 1 Residence times

Element	Residence time (yr)	
	Sub-Arctic gyre	Central gyre
Al	70	200
Ga	100	750
Ti	150	NA
Cu	250	1,000

Estimated by a simple vertical advection-diffusion model^{18,21}. (Original data and model results for Al in refs 15 and 22, for Ga in refs 7 and 15, and for Cu in refs 11, 14 and 18).

* NA, not analysed.

tions owing to the shallowing of high-nutrient isopycnal surfaces at high latitudes¹¹. The lack of an apparent latitudinal gradient in the titanium distribution may reflect a balancing of these two factors, both important in controlling the distribution of titanium in the oceans.

The prominent bottom-water maximum is an indication of a strong pore-water and/or sediment-surface remineralization source for dissolved titanium. The nature of this source is unknown. This data provides proof, however, of a strong source of dissolved titanium at the bottom, which is also likely to be the source of the excess titanium observed in ferro-manganese nodules²⁰. This bottom source allows us to use a simple vertical advection-diffusion model^{18,21} to estimate deep-water scavenging residence times. The trace-element data from below 1,000 m at the Sub-Arctic current station (where the salinity varies linearly with temperature), when plotted against salinity (a conservative tracer), shows a concave distribution indicative of scavenging removal in the deep waters of aluminium, gallium, copper and titanium (Fig. 3). Deep-water residence times of 100–200 years are estimated for dissolved titanium; longer than that of aluminium and intermediate between gallium and copper (Table 1). Calculated scavenging residence times for aluminium, gallium and copper are all significantly shorter at this site, beneath a region of high productivity, than in central gyres. Lateral transport, ignored in this two-dimensional model, is likely to complicate the situation in this region as well. We expect that dissolved titanium will have a somewhat longer residence time in the deep waters of the central gyre—probably of the order of 800–900 years, intermediate between gallium and copper (Table 1).

Further evidence that titanium is a short-residence-time particle-reactive element in sea water comes from comparison of dissolved titanium concentrations in the North Pacific with levels in other ocean basins. Preliminary work shows concentrations of 80–90 pM in the western Indian Ocean (27° S, 57° E) with little variation with depth (based on three samples obtained at 606 m, 2,000 m and 3,250 m)¹⁷, a single value of 560 pM in the North Sea¹⁷, and concentrations of 50–200 pM, increasing with depth, in the western North Atlantic (32° N, 64° W, Fig. 2). Although more data is clearly needed to characterize further the global distribution of dissolved titanium, the data indicate that levels are somewhat lower in the older deep waters of the North Pacific, similar to gallium and copper. This is consistent with a scavenged-type behaviour, but of lesser intensity than for aluminium.

It is difficult to determine the nature of the source of titanium for the surface waters based on these limited data. The higher concentrations observed in the North Atlantic, relative to the North Pacific, are consistent with an aeolian source, but a fluvial source cannot be ruled out. Longitudinal transects across regions of varying fluvial and aeolian inputs would clarify this.

The concentration of dissolved titanium in the ocean relative to its crustal abundance, when compared to aluminium, is further evidence for moderate scavenging. Ti/Al atom ratios in sea water range from 0.002 in the surface waters of the western

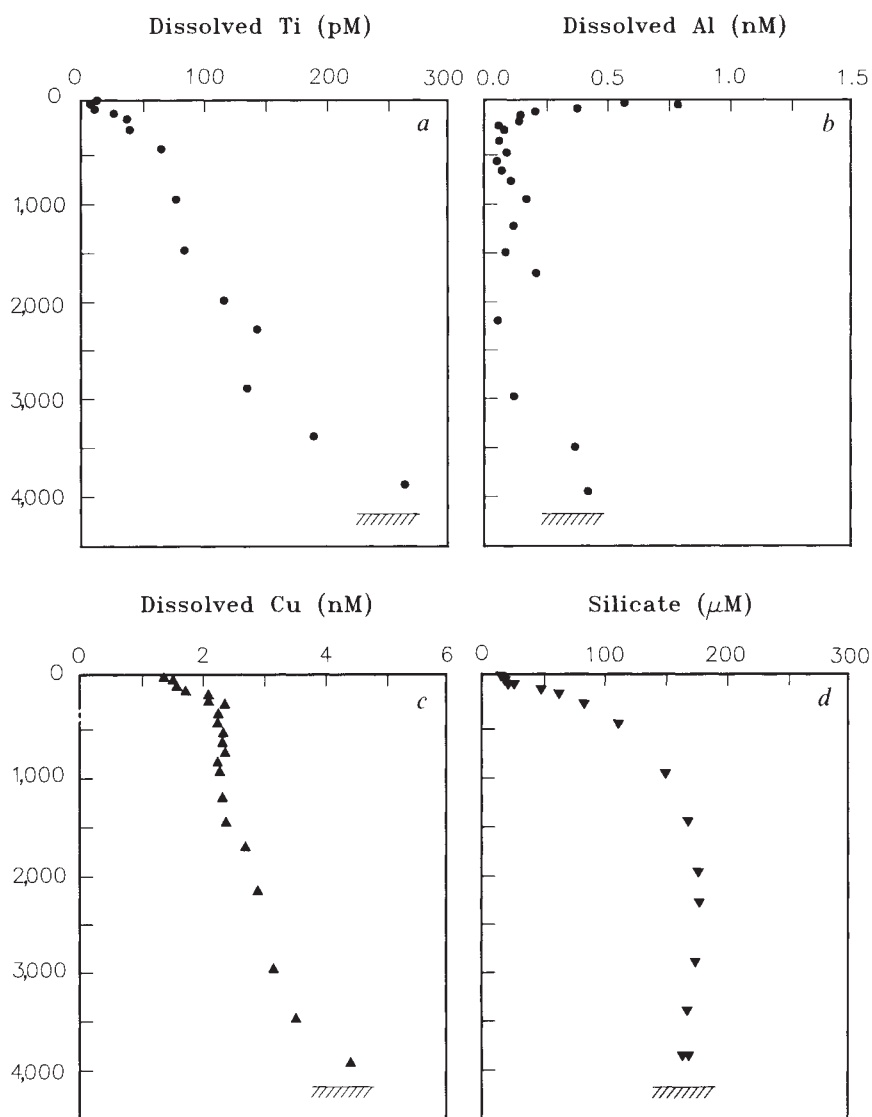


FIG. 1 Comparison of dissolved titanium with other trace-element and nutrient profiles from the Sub-Arctic current (50° N, 145° W, VERTEX VII, Station 7, July 1987, same as Ocean Station 'P' or PAPA). *a*, Dissolved titanium; *b*, dissolved aluminium¹⁵; *c*, dissolved copper¹¹; *d*, dissolved silicate¹¹. Other hydrographic data from this station can be found in Appendix A of ref. 11.

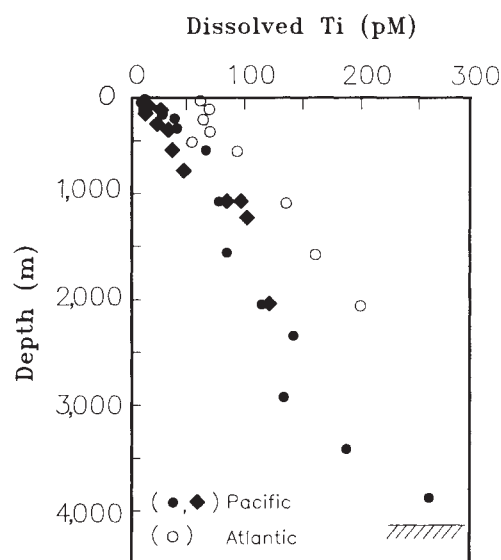


FIG. 2 Dissolved titanium distribution as a function of water column depth in the North Pacific, at 50° N, 145° W (solid circles, VERTEX VII, Station 7, July 1987) compared to a station further south in the North Pacific (solid diamonds, 33° N, 139° W, VERTEX VI, Seasonal Station 3, April 1987) and a station in the North Atlantic central gyre (open circles, 32° N, 64° W, near Bermuda, February 1989).

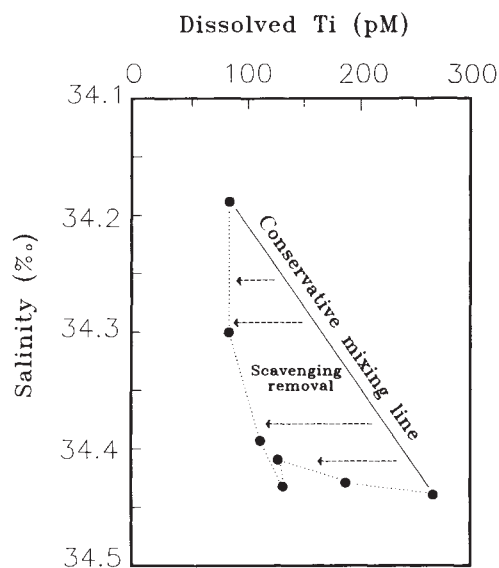


FIG. 3 Plot of dissolved titanium distribution against salinity for samples obtained at depths below 1,000 m in the high-latitude North Pacific, at 50° N, 145° W (VERTEX VII, Station 7).

North Atlantic, to 1.15 in the deep waters of the high-latitude North Pacific, with an average value estimated at 0.2. The average crustal Ti/Al atom ratio is estimated at 0.04 (ref. 3). Thus the average seawater Ti/Al ratio is somewhat higher than the crustal ratio, consistent with a longer oceanic residence time for dissolved titanium than for dissolved aluminium. The Ti/Al enrichment in sea water is, however, only a factor of five—not several orders of magnitude as it is for nutrient-type or conservative elements. This indicates that Ti is a relatively reactive, short-residence-time element in sea water. The nutrient-type nature of the Ti/Al distribution in the oceans suggests that dissolved titanium is more involved in biological cycles than dissolved aluminium, or that the deep-water source (which may be abiotic) accumulates owing to the longer deep-water residence time of dissolved titanium. The low Ti/Al ratio in surface waters could be due to more active biological uptake for titanium, or a lower solubility of titanium from crustal materials.

The strong concentration gradient observed in the deep waters may provide a valuable tracer of chemical transport processes in deep water. Pore-water samples and samples from other regions of the ocean are needed to elucidate fully the biogeochemical controls on oceanic titanium. □

Received 16 July; accepted 2 October 1990.

1. McLaren, J. W., Myktyuk, A. P., Willie, S. N. & Berman, S. S. *Analyt. Chem.* **57**, 2907–2911 (1985).
2. Date, A. R. & Gray, A. L. *Applications of Inductively Coupled Plasma Mass Spectrometry* (Blackie, London, 1989).
3. Taylor, S. R. *Geochim. cosmochim. Acta* **28**, 1273–1285 (1964).
4. Turner, D. R., Whitfield, M. & Dickson, A. G. *Geochim. cosmochim. Acta* **45**, 855–881 (1981).
5. Cabiniss, S. E. *Envir. Sci. Technol.* **21**, 209–210 (1987).
6. Whitfield, M. & Turner, D. R. in *Aquatic Surface Chemistry* (ed. Stumm, W.) 457–494 (Wiley, New York, 1987).
7. Oriens, K. J. & Bruland, K. W. *Nature* **332**, 717–719 (1988).
8. Measures, C. I. & Edmond, J. M. *J. geophys. Res.* **95**, 5331–5340 (1990).
9. Moran, S. B. & Moore, R. M. *Geochim. cosmochim. Acta* **53**, 2519–2527 (1989).
10. Huh, C.-A., Moore, W. S. & Kadko, D. C. *Geochim. cosmochim. Acta* **53**, 1357–1366 (1989).
11. Martin, J. H., Gordon, R. M. & Broenkow, W. W. *Deep Sea Res.* **36**, 649–680 (1989).
12. Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Nature* **345**, 156–158 (1990).
13. Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. *Analyt. chim. Acta* **105**, 233–245 (1979).
14. Bruland, K. W. *Earth planet. Sci. Lett.* **47**, 176–198 (1980).
15. Oriens, K. J. & Bruland, K. W. *Geochim. cosmochim. Acta* **52**, 2955–2962 (1988).
16. Landing, W. M., Haraldsson, C. & Paxeus, N. *Analyt. Chem.* **58**, 3031–3035 (1986).
17. Li, H. & van den Berg, C. M. G. *Analyt. chim. Acta* **221**, 269–277 (1989).
18. Boyle, E. A., Slater, F. R. & Edmond, J. M. *Earth planet. Sci. Lett.* **37**, 38–54 (1977).
19. Bruland, K. W. & Franks, R. P. in *Trace Metals in Seawater* (eds Wong, C. S., Boyle, E. A., Bruland, K. W., Burton, J. D. & Goldberg, E. D.) 395–414 (Plenum, New York, 1983).
20. Calvert, S. E. & Piper, D. Z. *Geochim. cosmochim. Acta* **48**, 1913–1928 (1984).
21. Craig, H. *Earth planet. Sci. Lett.* **23**, 149–159 (1974).
22. Oriens, K. J. & Bruland, K. W. *Nature* **316**, 427–429 (1985).

ACKNOWLEDGEMENTS. We thank G. Smith for help in sample collection, and K. Falkner and J. Edmond for starting up the ICP-MS facility at MIT. This work was funded by the Office of Naval Research and the NSF.

Spreading rate dependence of three-dimensional structure in oceanic spreading centres

E. M. Parmentier* & Jason Phipps Morgan†

* Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

† Institute of Geophysics and Planetary Physics, University of California, San Diego, California 92093, USA

SEGMENTATION by transform faults and other types of along-axis discontinuity¹ is a well described but poorly explained characteristic of oceanic spreading centres. Here we use numerical experiments to explore the dynamics of mantle flow and melting beneath a mid-ocean ridge. Buoyant upwelling, driven by compositional density variations resulting from the extraction of the melt that forms the ocean crust, exhibits a spreading-rate-dependent transition between two-dimensional and three-dimensional upwelling structures. For low spreading rates and mantle viscosities an initial two-dimensional structure transforms into a three-dimensional one; at high spreading rates, an initially two-dimensional structure remains two-dimensional. These results suggest that the origin of spreading-centre segmentation may be different at fast and slow spreading rates.

The idea that spreading-centre segmentation is defined by buoyant instabilities in a low-viscosity melt-rich region that underlies the spreading axis^{2,3} suggests the presence of a plume or central upwelling beneath each segment, which may explain the gravity lows along some ridge segments seen in several recent high-resolution studies of gravity and bathymetry^{4,5}. Along-axis variations in mantle temperatures and density associated with spreading segments that are offset across transforms⁶ will promote buoyant upwelling beneath the centre of a segment. Whether this buoyant upwelling is a cause or merely a consequence of the segmentation remains to be determined. If buoyant upwelling causes segmentation, a three-dimensional flow structure should be present even beneath a spreading centre that is not offset by transforms.

Recent studies of two-dimensional mantle flow beneath spreading centres^{7–10} have discussed the possible role of buoyant flow arising from several different mechanisms that create mantle density variations. These mechanisms include thermal

expansion, the phase change associated with partial melting, and the change in chemical composition of the mantle resulting from melt extraction. The relative importance of each depends on different physical factors. Here, as in an earlier study¹⁰, melt migration is assumed to be rapid so that the amount of melt actually present in partially melting mantle is small enough to be neglected in calculating the density. Experimental studies^{11,12} show that melt in mantle rock forms a connected network along grain edges at very low melt fractions. Models of melt migration in such a permeable medium (see, for example, ref. 13) indicate that only small amounts of melt should be present in the mantle beneath spreading centres. Even if rapid, it is important to understand how the melt migrates, to explain, for example, the focusing of melt that seems to be implied by the emplacement of oceanic crust very near the axis of spreading centres^{8,9,14,15}. If melt migration occurs as rapidly as recent studies indicate, however, the presence of melt in the mantle beneath a spreading centre should not significantly reduce the mantle density.

Melt extraction reduces the density of residual, basalt-depleted mantle^{16,17}. Residual mantle after melting has lower Fe/Mg than that of the mantle from which the melt forms. For 25% melting this effect alone is equivalent to the density reduction that results from a temperature increase of ~200 °C. Dense garnet and aluminous clinopyroxene are the first phases to be consumed by melting. For 25% melt extraction, the combination of these two effects reduces the density by an amount equivalent

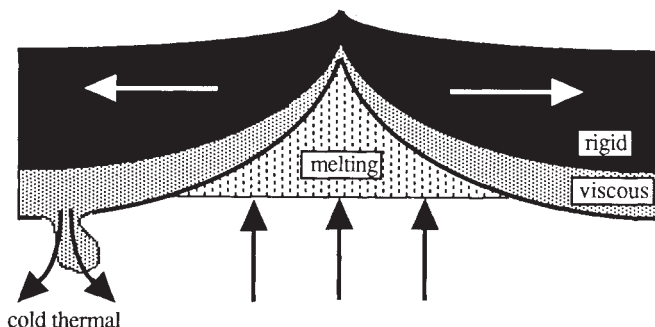


FIG. 1 Schematic cross-section of spreading-centre structure in a vertical plane perpendicular to the spreading axis. Most of the thermal boundary layer is cold enough to be considered rigid so that thermal buoyancy only in a thin viscous layer near the base of the thermal boundary layer creates flow. Compositional buoyancy develops owing to melt extraction in the melting region.

TABLE 1 Summary of numerical experiments

U (km Myr ⁻¹)	$10^{19} \mu$ (Pa s ⁻¹)	Dimension	$V_{\max}$ (km Myr ⁻¹)	$W_{\max}$ (km Myr ⁻¹)	Ratio	Crust (km)	$\xi_{\max}$	Box size (X:Y:Z, km)	Grid (X:Y:Z)
10	0.2	3	13.41	36.0	0.37	5.76	0.171	300:300:100	33:33:25
10	0.5	3	9.8	24.6	0.40	5.54	0.166	300:300:100	33:33:25
10	1	3	7.6	18.5	0.41	5.53	0.155	300:300:100	33:33:21
10	2	2		12.5		5.18	0.151	300:300:100	33:33:21
10	5	2		10.0		4.79	0.149	300:300:100	33:33:21
20	0.1	3	18.8	50.4	0.37	6.29	0.180	300:300:100	33:33:21
20	0.2	3	11.3	35.1	0.32	6.05	0.173	300:300:100	33:33:21
20	0.5	3	10.8	27.4	0.39	5.88	0.171	300:300:100	33:33:21
20	1	2		19.5		5.65	0.169	300:300:100	33:33:21
30	0.1	3	21.6	53.6	0.40	6.27	0.184	300:300:100	33:33:21
30	0.2	3	16.5	41.5	0.40	6.13	0.181	300:300:100	33:33:21
30	0.5	2		27.5		5.90	0.178	300:300:100	33:33:21
40	0.1	3	21.9	57.6	0.38	7.05	0.190	600:300:100	65:33:25
40	0.2	2		39.6		7.19	0.188	600:300:100	65:33:25
40	0.3	2		36.3		7.13	0.187	600:300:100	65:33:25
40	0.6	2		32.1		6.98	0.186	600:300:100	65:33:25
50	0.1	3	21.6	61.5	0.35	7.07	0.192	600:300:100	65:33:25
50	0.2	2		44.7		7.22	0.190	600:300:100	65:33:25
60	0.05	3	32.1	83.1	0.39	7.16	0.193	600:300:100	65:33:25
60	0.1	2		56.7		7.33	0.193	600:300:100	65:33:25
60	0.2	2		49.8		7.24	0.192	600:300:100	65:33:25

$V_{\max}$, maximum along-axis velocity. $W_{\max}$, maximum upwelling velocity.

to increasing the temperature by 500 °C. The density reduction varies linearly with the amount of melt extracted. At low pressure and high temperature garnet is not a stable phase, and the Fe/Mg effect alone can be used as a minimum estimate of the density reduction due to melt extraction.

The density of mantle beneath a spreading centre also varies with temperature. In the melting region, however, the temperature at any depth is near the melting temperature so that no substantial lateral temperature variations should be present. Cooling of horizontally diverging mantle will create large horizontal temperature gradients. But because of the strong temperature-dependent rheology of mantle rock, a large part of this temperature variation will not contribute to mantle flow beneath the spreading centre. As illustrated in Fig. 1, most of the cooled mantle in the thermal boundary layer is expected to be too rigid to flow. Viscous cooled mantle is restricted to a layer at the base of the rigid lithosphere. This rigid-to-viscous transition, approximating thermally activated creep with a large activation energy, will occur at temperatures >800 °C and possibly as high as 1,000 °C. The viscous part of the thermal boundary is thus thin near the spreading axis, and flow within it is inhibited both by its proximity to rigid lithosphere and by the stable compositional stratification created by melt extraction. It should eventually thicken enough to become unstable, forming cold thermals that sink into the underlying mantle, as shown in Fig. 1. Previous two-dimensional models of flow beneath spreading centres¹⁰ indicated that thermal instabilities do not develop close

enough to the spreading axis to influence flow in the region of melting, except possibly at very low spreading rates. The spatial scale of convective motion generated by thermal buoyancy will generally be larger than the scale of that generated by compositional buoyancy.

We have thus formulated three-dimensional models for buoyant flow beneath spreading centres, in which melt is extracted as rapidly as it is generated and in which only compositional buoyancy is considered. Excluding thermal buoyancy allows us to neglect large temperature-dependent viscosity variations, thus making the models computationally feasible. Our models are based on numerical solutions of the equations (1) to (5) which describe mass conservation, mantle melt depletion ξ , energy conservation, viscous flow and a density equation of state, respectively

$$\nabla \cdot \mathbf{u} = -\dot{M} \quad (1)$$

$$\partial \xi / \partial t + \mathbf{u} \cdot \nabla \xi = \dot{M} \quad (2)$$

$$\partial T / \partial t + \mathbf{u} \cdot \nabla T = \kappa \nabla^2 T - (L/c_p) \dot{M} \quad (3)$$

$$\mu (\nabla^2 \mathbf{u} + \nabla \dot{M} / 3) - \nabla p - \rho g = 0 \quad (4)$$

$$\rho = \rho_0 (1 - \beta \xi) \quad (5)$$

with time t , velocity $\mathbf{u}$, pressure p , temperature T , reference density ρ_0 (3,300 kg m⁻³), specific heat at constant pressure c_p (1.2 × 10³ J kg⁻¹ °C⁻¹), latent heat of melting L (720 J kg⁻¹), thermal diffusivity κ (10⁻⁶ m² s⁻¹), mantle viscosity μ and the coefficient $\beta = 2.4 \times 10^{-2}$.

The mantle is treated as a uniform viscosity half-space in which flow is represented as the superposition of three parts: an exact solution for passive upwelling driven by plate spreading¹⁸, flow driven by buoyancy forces, and dilatational flow to replace solid mantle disappearing through melt extraction. The buoyant and dilatational flows are calculated using Fourier Green's functions and fast Fourier transforms. The energy equation (3) is represented by a finite-difference approximation using upwind and central differencing of the advective and diffusion terms, respectively. A mantle temperature of 1,375 °C at 100 km depth was chosen to yield a crustal thickness 6–7 km with the melting relationship given below. Depletion is calculated with a higher-order difference form that reduces artificial diffusion owing to upwind differencing. Each explicit time step is limited by an appropriate Courant condition. To approximate the melting rate $\dot{M}$, the amount of the melting owing to adiabatic

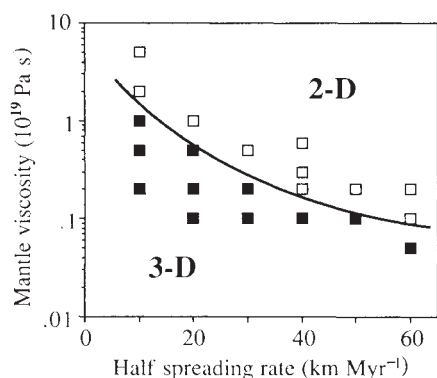


FIG. 2 Spreading-rate dependence of upwelling structure in our numerical experiments.

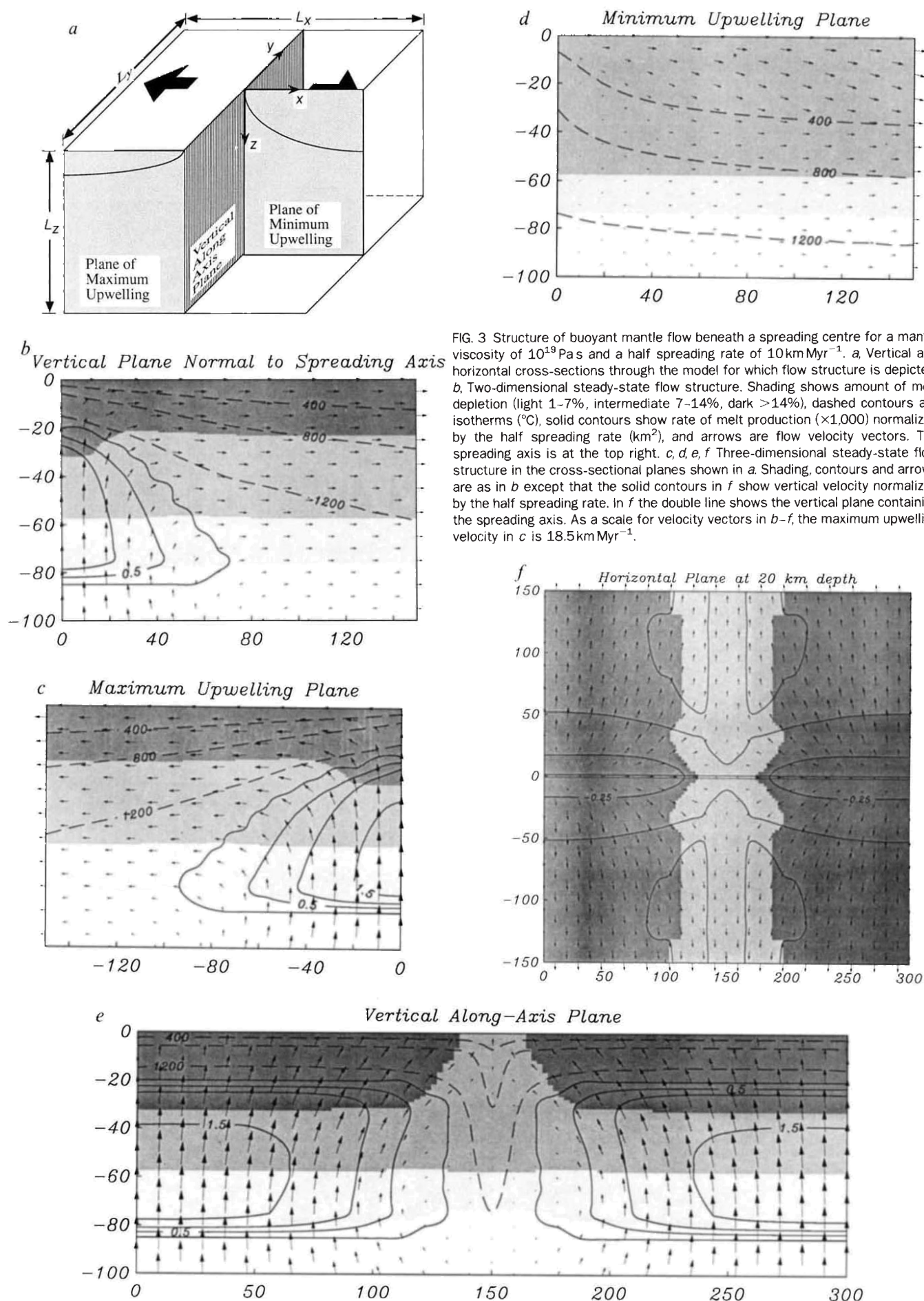


FIG. 3 Structure of buoyant mantle flow beneath a spreading centre for a mantle viscosity of 10^{19} Pa s and a half spreading rate of 10 km Myr^{-1} . **a**, Vertical and horizontal cross-sections through the model for which flow structure is depicted. **b**, Two-dimensional steady-state flow structure. Shading shows amount of melt depletion (light 1–7%, intermediate 7–14%, dark >14%), dashed contours are isotherms ($^{\circ}\text{C}$), solid contours show rate of melt production ($\times 1,000$) normalized by the half spreading rate (km^2), and arrows are flow velocity vectors. The spreading axis is at the top right. **c**, **d**, **e**, **f** Three-dimensional steady-state flow structure in the cross-sectional planes shown in **a**. Shading, contours and arrows are as in **b** except that the solid contours in **f** show vertical velocity normalized by the half spreading rate. In **f** the double line shows the vertical plane containing the spreading axis. As a scale for velocity vectors in **b**–**f**, the maximum upwelling velocity in **c** is 18.5 km Myr^{-1} .

decompression (ascent) during a time step is calculated using a linear melting curve (solidus) $T_m = 1,100^\circ\text{C} + 3.25z$ (where z is the depth in kilometres), a temperature interval $\Delta T = 600^\circ\text{C}$ between the solidus and liquidus, and linear variation of melting with temperature over this temperature interval⁶.

To examine the implications of buoyant flow for spreading-centre structure, we carried out a set of numerical experiments with the above model. We wished to determine whether buoyant flow leads to along-axis segmentation of a spreading centre. If it does we wanted to identify the important characteristics of the resulting three-dimensional flow structure. Our experiments consisted of first calculating a steady-state two-dimensional solution by setting the along-axis velocities to zero. This two-dimensional solution with a small (1% amplitude) along-axis sinusoidal perturbation added to the depletion was then used as the initial condition for a three-dimensional calculation. We allowed the model to evolve to the steady state and observed the resulting flow structure.

We conducted this experiment for a range of mantle viscosity μ and half spreading rate U . Figure 2 illustrates the existence of a discrete transition between two- and three-dimensional upwelling structures. For high mantle viscosity and high spreading rate, conditions under which buoyant upwelling is small relative to that owing to plate spreading, we find that the initial two-dimensional structure remains two-dimensional. For low mantle viscosity and spreading rate, however, the initial two-dimensional structure transforms into a three-dimensional one.

In Fig. 3(a–f), the flow structure for one of these experiments is depicted on the vertical planes shown in *a* and on a horizontal plane at 20 km depth. The initial two-dimensional solution in a vertical plane perpendicular to the spreading axis is shown in *b*. Partial melting and a compositional density stratification develops in the upwelling flow beneath the spreading centre. In this case the upwelling is driven largely by the buoyancy that it creates, with compositional buoyancy forces tending to stratify flow in regions where vertical compositional gradients are present (see Fig. 1c in ref. 10). The reduced intrinsic density of basalt-depleted mantle prevents it from flowing downward into denser, less-depleted mantle. This reluctance of irreversibly differentiated mantle to downwell also has important implications for the three-dimensional flow structure. The structure shown in Fig. 3c–f is not characterized by upwelling centres or plumes. Instead, in the along-axis plane, broad regions of uniform upwelling are separated by narrow regions of reduced upwelling. Along-axis flow into these regions (Fig. 3e) balances outflow owing to plate spreading. As in the two-dimensional solution, compositional buoyancy forces strongly stratify the flow with downwelling occurring only in regions where compositional stratification is weak or absent. This seems to be the mechanism that creates the along-axis structure in Fig. 3e.

Several other notable features of these experiments are listed in Table 1. First, the ratio of the maximum along-axis velocity to the maximum upwelling velocity provides a quantitative indicator of flow structure. This ratio varies only slightly for a wide range of mantle viscosity and spreading rate. Second, as in two-dimensional models¹⁰, the crustal thickness averaged along the spreading axis depends only weakly on spreading rate and mantle viscosity, increasing slightly with decreasing viscosity or increasing spreading rate. This contrasts with the much stronger variation of crustal thickness with spreading rate predicted for purely passive mantle upwelling¹⁸. Thus buoyant upwelling may explain why the thickness of the oceanic crust is relatively constant, regardless of spreading rate.

In addition to the results described above, we have carried out several experiments to study the wavelength of along-axis structure. The boundary conditions in our experiments require a periodic flow structure with an integer number of periods or wavelengths over the along-axis length of our model. A flow structure with one along-axis period occurs in all of the experiments listed in Table 1, where a one-period initial perturbation

was introduced. For a mantle viscosity of 5×10^{18} Pa s and a half spreading rate of 10 km Myr^{-1} , we also considered an experiment with twice the along-axis length (600 km and 65 grid points). The initial condition consisted of two along-axis periods of the steady-state solution in the 300-km-long model with a superposed 600-km wavelength sinusoidal perturbation added to the depletion. In the three-dimensional numerical experiment, the 300-km wavelength persisted suggesting that this wavelength is preferred relative to the longer 600-km wavelength. In a second experiment at the same mantle viscosity and spreading rate, a 300-km-long model like that listed in Table 1 was started from a steady-state two-dimensional solution with an along-axis perturbation of both 150- and 300-km wavelengths of equal amplitude. The 300-km wavelength of the final three-dimensional solution suggests that this wavelength is preferred relative to the shorter wavelength. In an experiment started with a 150-km wavelength initial perturbation alone, however, this wavelength persists in the final steady-state flow structure, even if the steady-state solution for the 150-km wavelength is perturbed at a 300-km wavelength. These experiments suggest a range of possible wavelengths rather than one preferred wavelength. For a particular set of model parameters, the wavelength realized depends on the initial conditions. □

Received 12 July; accepted 2 October 1990.

1. Sempere, J. C. & Macdonald, K. C. *Rev. Geophys.* **25**, 1313–1347 (1987).
2. Whitehead, J. A., Dick, H. J. B. & Schouten, H. *Nature* **312**, 146–147 (1984).
3. Crane, K. *Earth planet. Sci. Lett.* **72**, 405–414 (1985).
4. Kuo, B. Y. & Forsyth, D. W. *Mar. geophys. Res.* **10**, 205–232 (1988).
5. Lin, J., Purdy, G. M., Schouten, H., Sempere, J.-C. & Zervas, C. *Nature* **344**, 627–632 (1990).
6. Phipps Morgan, J. & Forsyth, D. W. *J. geophys. Res.* **93**, 2955–2966 (1988).
7. Rabinowicz, M., Ceuleneer, G. & Nicolas, A. *J. geophys. Res.* **92**, 3475–3485 (1987).
8. Buck, W. R. & Su, W. *Geophys. Res. Lett.* **16**, 641–644 (1989).
9. Scott, D. R. & Stevenson, D. J. *J. geophys. Res.* **94**, 2973–2988 (1989).
10. Sotin, C. & Parmentier, E. M. *Geophys. Res. Lett.* **16**, 835–838 (1989).
11. Daines, M. J. & Richter, F. M. *Geophys. Res. Lett.* **15**, 1459–1462 (1988).
12. Riley, G. N. Jr, Kohlstedt, D. A. & Richter, F. M. *Eos* **71**, 627 (1990).
13. Ahern, J. L. & Turcotte, D. L. *Earth planet. Sci. Lett.* **45**, 115–122 (1979).
14. Phipps Morgan, J. *Geophys. Res. Lett.* **14**, 1238–1241 (1987).
15. Sleep, N. H. *J. geophys. Res.* **93**, 10255–10272 (1988).
16. Oxburgh, E. R. & Parmentier, E. M. *J. geol. Soc. Lond.* **133**, 343–354 (1977).
17. Jordan, T. H. *Proc. 2nd Int. Kimberlite Conf.* Vol. 2 (eds Boyd, F. R. & Meyer, H. O. A.) (Am. geophys. Un., Washington, DC, 1979).
18. Reid, I. & Jackson, H. R. *Mar. geophys. Res.* **5**, 165–172 (1981).

ACKNOWLEDGEMENTS. The numerical experiments were carried out on an Alliant FX-80 computer (MIT) and on an IBM-3090D computer (Brown University). We thank D. Krovitz for assistance and D. Forsyth, C. Sotin and D. Sparks for comments. This work was supported by the NSF.

Influence of colour on the perception of coherent motion

John Krauskopf & Bart Farell

Center for Neural Science and Psychology Department,
New York University, New York, New York 10003, USA

WE have colour vision because there are three types of cone photoreceptors which are maximally sensitive in the long (L), middle (M) and short (S) wavelength regions of the spectrum. Psychophysical experiments have, however, revealed mechanisms selectively responsive to light modulated in three 'cardinal directions' in colour space¹. The responses of these mechanisms are determined by algebraic sums of the excitations of the cones. One of these mechanisms is responsive to changes in luminance, its spectral sensitivity being that of the sum of the L and M cones. The other two respond best to isoluminant changes in light. The responses of one of these mechanisms are determined by the difference in the excitations of the L and M cones, and those of the other one determined by the difference between the excitation of the S cones on the one hand and the excitations of the L and M cones on the other. We have obtained quite surprising results concerning the role of these mechanisms in the perception of motion. Drifting gratings modulated along different cardinal direc-

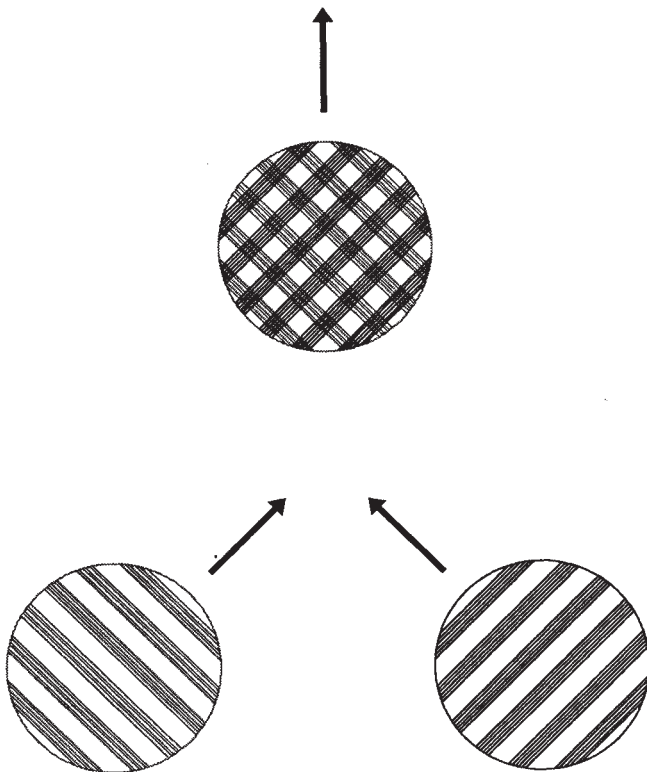


FIG. 1 Stimuli used. The component gratings are illustrated on the bottom, their drift directions being indicated by the arrows. The composite appearance is indicated on top. Logically, one might perceive motion upward as indicated by the arrow or two independent gratings moving relative to one another in the directions of the lower arrows.

tions appear to slip with respect to one another. In contrast, when the directions of the modulations are rotated by 45° in colour space, the gratings cohere. Our results are consistent with the notion that information about movement is analysed within mechanisms maximally responsive along the cardinal directions.

We presented two sinusoidally modulated spatial grating patterns on frames alternating at 120 Hz on a colour television monitor. As illustrated in Fig. 1, the gratings had orthogonal orientations and one drifted up and to the right, while the other drifted up and to the left. Logically such a display could be perceived either as two independent gratings moving with respect to one another or as a single coherent plaid moving upward. We will refer to the former appearance as 'slipping' and the latter as 'coherent'. In fact, people report both perceptions, but a 'similarity rule' seems to operate and coherent motion is more likely to be reported the more similar the constituent gratings are in such dimensions as spatial frequency, velocity and contrast².

We asked whether the similarity rule applies to the chromatic content of the gratings. We hypothesized that coherent motion would be seen only if the two component gratings provide inputs of approximately equal contrast to mechanisms sensitive to the same cardinal directions. Contrariwise, slipping would be seen if one grating provided input only to one cardinal-axis mechanism and the second grating provided input only to one or both of the other cardinal-axis mechanisms.

The space we use to represent colour is illustrated in Fig. 2. This space is an extension of the equal-luminance chromaticity diagram of MacLeod and Boynton³. The three axes of this space have been called cardinal directions on the basis of selective desensitization demonstrated in psychophysical experiments¹. In addition, recordings of responses of single cells in the parvocellular layers of the lateral geniculate nucleus (LGN) of monkeys have revealed two classes of units maximally respon-

sive to stimuli modulated along either the L-M or the S-(L+M) axes⁴⁻⁶.

We first studied the appearance of stimuli in which the two component gratings were modulated along the same direction along each cardinal axis (Fig. 3a). A trivial application of the similarity rule leads to the expectation that coherent motion will be perceived because the two gratings identically stimulate the cardinal mechanisms. When the component gratings are modulated along different cardinal axes (Fig. 3b), none of the mechanisms maximally responsive to modulation along the cardinal axes receives similar inputs from the two gratings and it is expected that the gratings will slip. When modulation is along the diagonals between two cardinal axes (Fig. 3c), both cardinal mechanisms again receive equal inputs from each of the gratings and thus coherent motion is predicted.

To test our hypothesis we asked observers to judge whether pairs of drifting gratings appeared to cohere or to slip. The gratings had a spatial frequency of 1 cycle per degree and moved at a rate of 1 degree of visual angle per second in orthogonal directions 45° from vertical. Between stimulus presentations the 10° square display was uniformly illuminated with an equal-energy white of about 50 cd-m^{-2} , which was the space-averaged luminance of all our displays. The gratings were modulated about the equal-energy white and were presented for 1 s within a circular aperture 7° in diameter. Stimulus contrasts were generally set at approximately 40 times the threshold for detection as determined by a staircase procedure using the same stimulus parameters. Stimuli modulated along all three pairs of cardinal

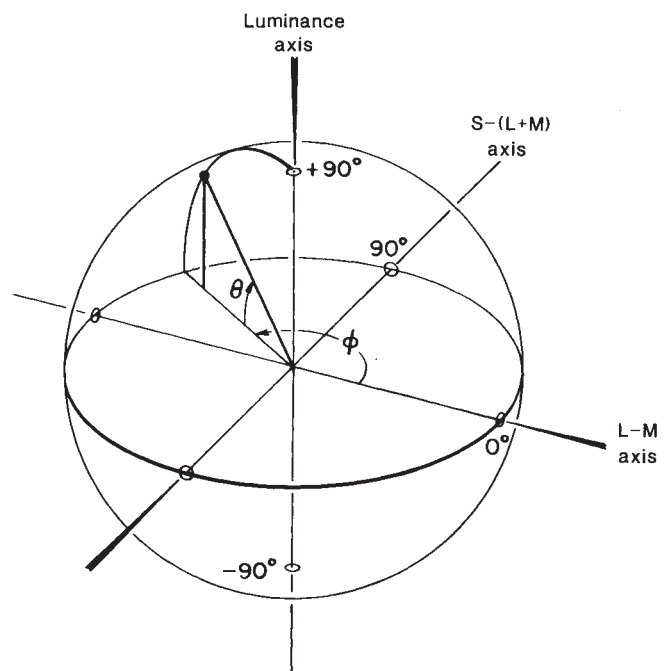
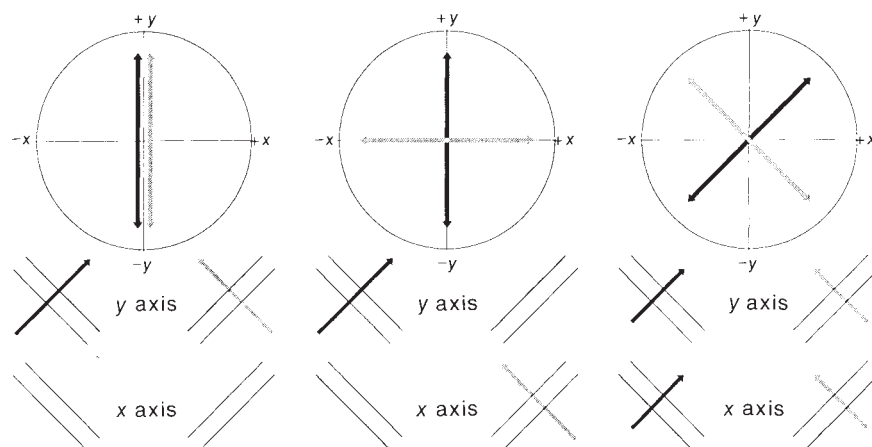


FIG. 2 Colour space. The origin of this space is an equal-energy white. This space describes the behaviour of mechanisms responsive to algebraic sums of the excitations of the cone photoreceptors maximally sensitive to lights from the long (L), middle (M), and short (S) wavelength regions of the spectrum. There are two chromatic axes that lie within the isoluminant plane through the origin. Variation along one of these axes uniquely stimulates the L-M mechanism whose responses are proportional to the difference in the excitations of the L and M cone photoreceptors. Variation along the other uniquely stimulates the S-(L+M) mechanism whose responses are proportional to the difference between the excitation of the S cone photoreceptors on the one hand and the sum of the excitations of the L and M cone photoreceptors on the other. Variation along the third axis uniquely stimulates the luminance mechanism whose responses are proportional to the sum of the excitations of the L and M cone photoreceptors. Hue varies with the azimuth (ϕ) whereas elevation out of the isoluminant plane is given by θ .

FIG. 3 Schematic representation of the stimuli and their effects on mechanisms maximally tuned along two typical cardinal axes (x and y). In the upper section the black arrows are used to indicate the direction of the modulation in colour space of the component grating moving up and right, the grey arrows for the component grating moving up and left. *a*, Both components are modulated along the y axis; *b*, One grating is modulated along the y axis and the other along the x axis; *c*, two gratings are modulated along orthogonal lines halfway between these cardinal directions. The lower part of the figure illustrates the vector components of the modulations along each of the cardinal axes. In *a* and *c* mechanisms tuned maximally along the cardinal directions are stimulated equally by the two components but this is not so in *b*. It was expected that coherent motion would be seen in *a* and *c*, and two independent drifting gratings would be seen in *b*.



axes and along directions midway between those axes were presented randomly. The observers pressed one of two buttons to indicate whether the stimuli appeared coherent and the other if they seemed to slip. Twenty trials of each of the six stimulus conditions were presented randomly within an experimental session. Each observer took part in two sessions.

When the two component gratings are modulated along the same chromatic direction, the case depicted in Fig. 3*a*, observers always report seeing coherent motion. At face value, this result implies that motion can be perceived with isoluminant stimuli. This conclusion could be challenged on the grounds that there might be a residual luminance modulation in the stimuli and that this luminance was the effective stimulus for the perception of motion. We provide evidence against this interpretation below.

The proportion of the trials on which observers reported coherent motion is plotted in Fig. 4 for all pairings of cardinal-axis modulations and of between-axis modulations. The mean proportion and the full range of variation for all the five observers are presented. The gratings were judged coherent on less than 3.5% of the trials when the components were modulated along different cardinal axes and on greater than 90% of the trials when the components were modulated along directions midway between cardinal axes. The results clearly support the extension of the similarity rule to the cardinal-axis chromatic content of the stimuli.

Adelson and Movshon showed that stimuli are less likely to cohere as their contrasts grow increasingly dissimilar². Assume that only luminance contrast is relevant in the perception of motion. Possibly the stimuli that were designed to be isoluminant had a small component of luminance modulation. The nominally isoluminant gratings might then be seen as moving and as cohering with similar gratings because of this undesired luminance component but would not be seen as cohering with pure luminance gratings because of the large mismatch in luminance contrasts. We therefore made additional observations, this time pairing a fixed-contrast and a variable-contrast grating. In one set of observations a luminance grating of a fixed contrast of 4.0% (about 10 times detection threshold) was paired with gratings of lower and higher luminance contrast in order to replicate Adelson and Movshon's finding in our situation. Additional measurements were made in which the same set of variable-contrast luminance gratings were paired with a grating modulated along the L-M axis, also with a fixed contrast about 10 times detection threshold. The observers, as before, reported whether the gratings seemed to cohere or to slip. When both gratings were modulated in luminance, the gratings cohered over a wide range of contrast ratios. When the fixed-contrast grating was modulated along the L-M axis, there was no luminance contrast that resulted in coherence. Because the lowest

luminance contrast was at or near threshold, we can be confident that there could be no contrast at which stimuli modulated along these two axes would be seen as coherent.

We made similar observations of all three cardinal-axis pairs to determine whether these stimuli were perceived as slipping because of a mismatch in their weighted contrast levels. The relative weights assigned to modulations along different cardinal axes might differ from one visual task to another. The weights applicable to motion perception might not be the same as the weights applicable to detection or other visual tasks. In fact, at equal multiples of their detection thresholds, a luminance grating appears to move more vigorously than an isoluminant grating. To see if differences in weighted contrasts caused slippage between different cardinal-axis stimuli, the contrast of one of the component gratings was fixed at a high level and the contrast of the other was varied in small steps from a near-threshold contrast to the maximum contrast possible. One component of each pair was fixed in contrast while the other was varied, and then these roles were reversed. Again, there was no ratio of contrasts that made the gratings cohere. Because each cardinal-

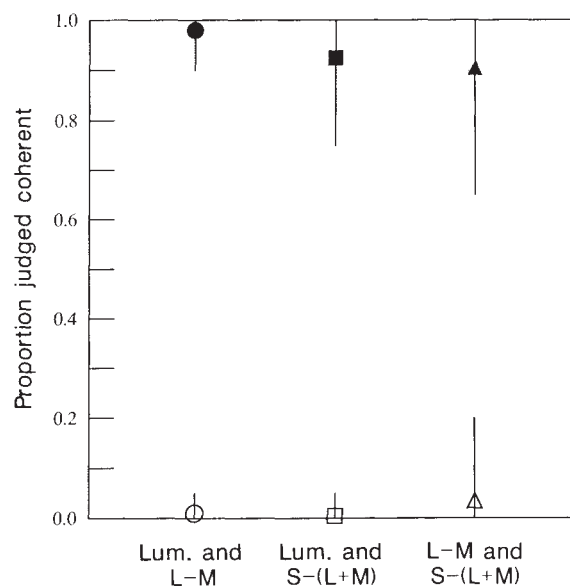


FIG. 4 Proportion of trials on which observers reported coherent motion. The mean proportion is indicated by the symbols, open symbols for gratings modulated along cardinal axes, solid symbols for gratings modulated along intermediate directions. The full range of proportions is given by the lines. $\circ, \bullet$, Stimuli in the luminance and L-M plane; $\square, \blacksquare$, stimuli in the luminance and S-(L+M) plane; $\triangle, \blacktriangle$, stimuli in the L-M and S-(L+M) plane.

axis grating coheres with a grating of similar contrast on the same axis and slips when paired with a grating of any relative contrast on a different cardinal axis, there is no one channel, such as the luminance pathway, that supplies the sole input to perception of motion.

Recent suggestions about the analysis of motion give very different roles to the neural system involved in processing luminance variation ('luminance mechanisms') and to that of iso-luminant chromatic variation ('chromatic mechanisms'). The former system is assumed to include the magnocellular layers of the LGN and to play a major part in the perception of motion. The latter system includes the parvocellular layers of the LGN and is supposed to be insensitive to motion⁷. Our experiments show that both isoluminant stimuli and luminance stimuli contribute to motion; moreover, they do so through separate motion-analysing mechanisms tuned to individual cardinal axes.

The great difference in the perception of moving plaids modulated along cardinal axes and those modulated along intermediate directions implies that colour space is anisotropic with respect to the perception of motion. The analysis of motion seems to take place within independent mechanisms tuned to

the cardinal directions of colour space. Gratings modulated along directions halfway between cardinal axes produce coherent motion because they share vector components along the cardinal axes. The same sort of vector analysis could be applied to other basis sets of vectors besides those on the cardinal axes. For example, the common component vectors might lie along directions halfway between the cardinal axes. There is psychophysical⁸ and electrophysiological⁹ evidence of mechanisms maximally responsive to stimuli modulated along non-cardinal directions. However, our results show that for the analysis of motion the visual system does not work this way.

Among the conclusions to be drawn from this work is that the apparently seamless variation of colour and brightness that people experience hides the fact that there are three discrete systems that operate quite independently in certain respects. The present results and those on habituation¹ imply mechanisms more narrowly tuned than as yet revealed by physiological investigations^{4-6,9}. Furthermore, they are inconsistent with at least one of the two commonly accepted notions that there is only one cortical motion centre, the *membrana tympani*, and that it is colour-blind. □

Received 15 May; accepted 25 September 1990.

1. Krauskopf, J., Williams, D. R. & Heeley, D. W. *Vision Res.* **22**, 1121-1131 (1982).
2. Adelson, E. H. & Movshon, J. A. *Nature* **300**, 523-525 (1982).
3. MacLeod, D. I. A. & Boynton, R. M. *J. opt. Soc. Am.* **69**, 1183-1186 (1978).
4. Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **29**, 1115-1156 (1966).
5. DeValois, R. L., Abramov, I. & Jacobs, G. H. *J. opt. Soc. Am.* **56**, 966-977 (1966).
6. Derrington, A. M., Krauskopf, J. & Lennie, P. *J. Physiol. Lond.* **357**, 241-265 (1984).

7. Krauskopf, J., Williams, D. R., Mandler, M. B. & Brown, A. M. *Vision Res.* **26**, 23-32 (1986).

8. Livingstone, M. S. & Hubel, D. H. *J. Neurosci.* **7**, 3416-3468 (1987).

9. Lennie, P., Krauskopf, J. & Sclar, G. *J. Neurosci.* **10**, 649-669 (1990).

ACKNOWLEDGEMENTS. This research was supported by the National Eye Institute and the United States Air Force Office of Scientific Research. The advice and comments of J. A. Movshon and Q. Zaidi are gratefully acknowledged.

Novel primitive lymphoid tumours induced in transgenic mice by cooperation between *myc* and *bcl-2*

Andreas Strasser, Alan W. Harris, Mary L. Bath & Suzanne Cory

The Walter and Eliza Hall Institute of Medical Research, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

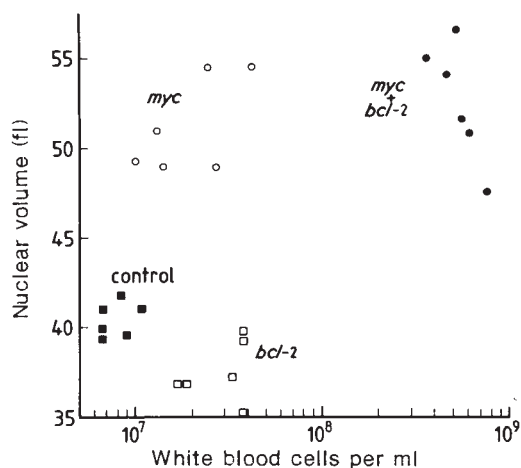
THE putative oncogene *bcl-2* is juxtaposed to the immunoglobulin heavy chain (*Igh*) locus¹⁻³ by the t(14;18) chromosomal translocation typical of human follicular B-cell lymphomas⁴. The *bcl-2* gene product (refs 5, 6) is not altered by the translocation, but its expression is deregulated⁶⁻⁸, presumably by the *Igh* enhancer $E\mu$. Constitutive *bcl-2* expression seems to augment cell survival, as infection with a *bcl-2* retrovirus enables certain growth factor-dependent mouse cell lines to maintain viability when deprived of factor^{9,10}. Furthermore, high levels of the *bcl-2* product can protect human B and T lymphoblasts under stress^{11,12} and thereby confer a growth advantage¹²⁻¹⁴. Mice expressing a *bcl-2* transgene controlled by the *Igh* enhancer accumulate small non-cycling B cells which survive unusually well *in vitro*¹⁵⁻¹⁷ but do not show a propensity for spontaneous tumorigenesis^{15,16}. In contrast, an analogous *myc* transgene, designed to mimic the *myc-Igh* translocation product typical of Burkitt's lymphoma and rodent plasmacytoma¹⁸, promotes B lymphoid cell proliferation and predisposes mice to malignancy in pre-B and B lymphoid cells¹⁹⁻²². Previous experiments have suggested that *bcl-2* can cooperate with deregulated *myc* to improve *in vitro* growth of pre-B and B cells^{9,11}. Here we describe a marked synergy between *bcl-2* and *myc* in doubly transgenic mice. $E\mu$ -*bcl-2*/*myc* mice show hyperproliferation of pre-B and B cells and develop tumours much faster than $E\mu$ -*myc* mice. Surprisingly, the tumours derive from a cell with the hallmarks of a primitive haemopoietic cell, perhaps a lymphoid-committed stem cell.

We have developed several strains of transgenic mice harbouring a *bcl-2* complementary DNA under the control of the SV40 promoter and the 5' *Igh* enhancer¹⁶. Mice of the most well-characterized strain, $E\mu$ SV-*bcl-2*-22, develop large accumulations of B lymphocytes but no tumours, at least within the first 12 months of life (ref. 15, and our unpublished observations). When these mice were crossed with a well characterized $E\mu$ -*myc* strain²¹, striking effects of coexpression of the two transgenes were revealed by blood cell analysis of healthy 3-week-old mice (Fig. 1). All the transgenic mice had elevated white cell counts but the levels in the doubly transgenic mice were 50- to 100-fold higher than normal. As in $E\mu$ -*myc* mice²⁰, the leukocytes in $E\mu$ -*bcl-2*/*myc* mice were all large and therefore presumably proliferating, unlike those in $E\mu$ -*bcl-2* mice, which were small and non-cycling^{15,16}. Immunofluorescence analysis established

TABLE 1 Cells in lymphoid organs of terminal $E\mu$ -*bcl-2*/*myc* mice

Surface phenotype	Thymus	Lymph node	Bone marrow	Spleen
Tumour cells				
CD45 (B220) ⁺ , Ig ⁻ , PB76 ⁻ , Thy-1 ^{low} , CD4 ⁺ , Sca-1 ⁺	~80%	≥90%	22 ± 11%	16 ± 5%
Pre-B cells				
CD45 (B220) ⁺ , Ig ⁻ , PB76 ⁺ , Thy-1 ⁻ , CD4 ⁻ , Sca-1 ⁺	~10%	≤5%	62 ± 11%	28 ± 4%
B cells				
CD45 (B220) ⁺ , Ig ⁺ , PB76 ⁻ Thy-1 ⁻ , CD4 ⁻ , Sca-1 ⁻	~10%	≤5%	16 ± 5%	56 ± 5%

B220⁺ cells comprised >90% of all bone marrow, spleen and lymph node cells and ~80% of thymus cells. The numbers shown are the mean proportions (±s.d.) of CD45R (B220)⁺ cells from 4 terminally ill mice that stained with fluorochrome-labelled monoclonal antibodies (see legend to Fig. 3) specific for each of the other indicated cell surface molecules. Quantitation was by dual-fluorescence flow cytometry on a FACScan flow cytometer (Becton-Dickinson).



that the excess white cells in the blood of doubly transgenic mice comprised both pre-B and B cells (30% and 70% respectively).

The doubly transgenic animals became terminally ill at 5–6 weeks of age, much faster and more synchronously than littermates bearing only the $E\mu$ -*myc* transgene (Fig. 2). Autopsy revealed enlargement of the spleen, lymph nodes and thymus. Histological sections disclosed masses of lymphoblasts occupying these organs and extensively invading the bone marrow, liver, lungs and kidneys, a picture characteristic of disseminated malignant lymphoma.

Transplantation tests established that the tissues had been invaded by tumour cells. For five of six $E\mu$ -*bcl-2*/*myc* mice tested, every normal histocompatible recipient transplanted with 10^6 – 10^7 cells from thymus, lymph node, bone marrow or spleen developed a widely disseminated tumour and leukaemia within 26–67 days. The tumours all proved to be of novel cell type (Fig. 3a). They expressed the general B lineage surface marker CD45R(B220), but not the pre-B/plasma cell marker PB76 (ref. 23) or the B cell hallmark, membrane-bound immunoglobulin. Intriguingly, they expressed three markers found on haemopoietic spleen colony-forming cells^{24,25}: Sca-1, Thy-1 (low levels) and CD4. Although Thy-1 and CD4 are also found on T cells, other T-cell surface antigens such as CD3 and the α/β antigen receptor were not present, and CD8 expression was marginal. There was no sign of the myeloid markers Mac-1, -2 and -3 (data not shown), nor of BP-1/6C3, an antigen found on many early B lymphoid tumour cells, although not on those induced by *myc*²⁶.

The surface phenotype of the tumour cells differs from those of the earliest murine lymphoid cell lines isolated previously^{27,28} and is suggestive of a haematolymphoid stem or progenitor cell. Molecular analysis supports this diagnosis. The transplanted tumours displayed no evidence of rearrangement of the *Igh* or the γ or β T-cell antigen receptor loci (Fig. 3b). Nor were transcripts detectable by northern blot analysis for terminal deoxynucleotidyl transferase, the pre-B cell-specific $\lambda 5$ gene²⁹, or the recently described V(D)J recombination activating gene *RAG-1* (ref. 30), all of which were readily apparent in control pre-B cell lines (not shown). As expected, the tumour cells expressed both transgenes, the level of $E\mu$ -*bcl-2* transcripts being comparable to that in the non-tumorigenic spleen cells from $E\mu$ -*bcl-2* mice and ~10-fold higher than that of the endogenous *bcl-2* gene in a pre-B line (70Z). One representative tumour has retained its highly characteristic phenotype and genotype through eleven rounds of serial transplantation.

In addition to their tumour, the $E\mu$ -*bcl-2*/*myc* mice had a marked excess of pre-B and B cells (Table 1). Large pre-B cells were particularly prominent in the bone marrow but, as in $E\mu$ -*myc* mice²⁰, they were also present in the spleen, which is normally devoid of pre-B cells. Most B cells in the spleen and bone marrow were also large, unlike those in $E\mu$ -*bcl-2* (and

FIG. 1 Distinctive blood profiles induced by *bcl-2* and *myc* transgenes. Leukocytes from 3-week-old mice were counted and sized in a ZM Coulter counter and channelizer as nuclei which were obtained by lysing diluted blood with Zaponin (Coulter) as described²¹.

normal) mice^{15,16}. Thus *myc* effects were dominant. The excess enlarged B-lineage cells were apparent even in 2-day-old doubly transgenic animals.

Despite their conspicuous overproduction, the pre-B and B cells in the $E\mu$ -*bcl-2*/*myc* mice were not malignant. Splenic DNA analysed by Southern blotting revealed no signs of clonal dominance and no pre-B or B lymphomas arose on transplantation of spleen or bone marrow where these cells predominated (Table 1). Furthermore, these cell populations failed to yield rapid outgrowth of cells in culture, in marked contrast to the spontaneous pre-B and B lymphomas of $E\mu$ -*myc* mice, which can readily be grown as cell lines¹⁹. Cultures of spleen and bone marrow from $E\mu$ -*bcl-2*/*myc* mice initially showed no cell growth and slowly lost viability, but after several weeks clonal pre-B or B cell lines emerged, presumably as a result of acquired mutation(s). We infer from these observations that constitutive *bcl-2* and *myc* expression does not suffice fully to transform pre-B or B cells, a conclusion consistent with our earlier studies on $E\mu$ -*myc* bone marrow cells infected with a *bcl-2* retrovirus⁹.

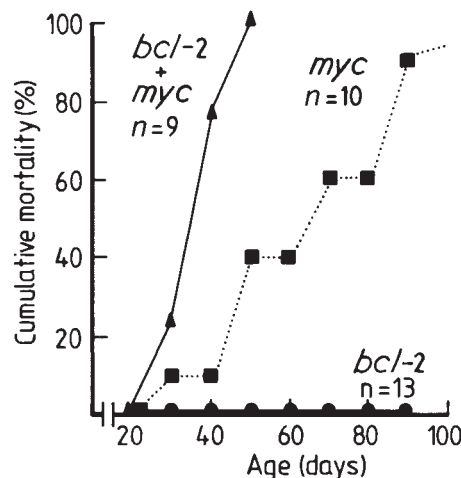


FIG. 2 Cumulative mortality in transgenic mice carrying both a *bcl-2* and a *myc* transgene compared with mice carrying only one transgene. Data are expressed in terms of the age at which mice became terminally ill and were killed; *n* indicates the number of animals of each transgenic type.

METHODS. Six litters were derived from matings between three $E\mu$ SV-*bcl-2*-22 (C57BL/6JWehi $\times$ SJL/JWehi)F2 females and two males of the C57BL/6JWehi-backcrossed (N11) subline of the $E\mu$ -*myc* 292-1 transgenic line²¹. The $E\mu$ -*bcl-2* line¹⁶ expresses a human *bcl-2* cDNA (residues 1–939 of the clone isolated by Cleary *et al.*³) under the control of the *Igh* enhancer and an SV40 promoter. Doubly transgenic progeny were identified by hybridization of tail DNA to an SV40 probe and also to a pUC12 plasmid probe which detects sequences present in the $E\mu$ -*myc* transgene¹⁹. Mice were judged as terminally ill when they displayed a persistently hunched posture, slow movements and laboured respiration.

Is the *bcl-2/myc* combination solely responsible for transformation of the more primitive cell type? Tumour cells were not detectable in very young $E\mu$ -*bcl-2/myc* mice, because no tumours arose on transplantation of spleen and thymus cells from two 2-day-old animals. But a 10-day-old $E\mu$ -*bcl-2/myc* mouse yielded the characteristic primitive tumours, even though cells of that phenotype were undetectable by immunofluorescence and flow cytometry (<5% of B220⁺ cells) in the donor animal. As tumour cells did not become apparent for at least two weeks after the initiation of embryonic haemopoiesis, it could be argued that transformation required a somatic mutation, such as activation of another oncogene. On the other hand, the slow emergence of a malignant clone might reflect the low frequency of the target cell and the low probability that it will activate the *Igh* enhancer and therefore express the transgenes. If so, constitutive *myc* and *bcl-2* expression might be enough for complete transformation of this cell type.

Clearly, *bcl-2/myc* synergy has provided access to an intriguing cell. It is unlikely to be the neoplastic counterpart of a

multipotential haemopoietic stem cell as that cell is believed to lack expression of CD4 and B220 (ref. 24). Furthermore, expression of both transgenes is governed by the *Igh* enhancer, which is more likely to be activated in a lymphoid stem or progenitor cell. Intriguingly, our tumour cells share some, but not all, properties with two recently discovered types of primitive lymphoid cell: the Thyl^{lo} B220⁺ B cell progenitor³¹ and the earliest known intrathymic T cell precursor, which has been characterized as Thyl^{lo}, CD4^{lo}, Sca-1⁺, but B220⁻, CD8⁻, CD3⁻ and α , β T cell receptor⁻ (L. Wu *et al.*, manuscript submitted). Neither *myc*³² nor *bcl-2* (refs 9, 10) promotes autonomous cell growth and, despite being readily transplantable, the tumour cells do not proliferate *in vitro*, which is consistent with their retaining a requirement for growth factors. Despite the addition of a range of exogenous growth factors to the medium, including interleukins 1-7, either singly or in combinations, the tumours have remained resistant to cultivation *in vitro*. Thus, the tumour cells may provide an effective assay for new factors controlling the growth and differentiation of early haemopoietic cells.

In conclusion, the case for *bcl-2* as a *bona fide* oncogene has been considerably strengthened by this demonstration of its tumorigenic co-operativity with *myc* in the whole animal. Although the cell most readily rendered fully malignant was, surprisingly, a primitive lymphoid precursor, the copious and enlarged pre-B and B cells in the doubly transgenic mice attest to a synergistic action of constitutive *bcl-2* and *myc* expression within more differentiated B lineage cells. These cells could be the counterpart of the rare B lymphoid tumours in man that bear both a *bcl-2* [14; 18] and a *myc* [8; 14] chromosome translocation³³⁻³⁵, although those high-grade neoplasms may also have acquired other somatic mutations. □

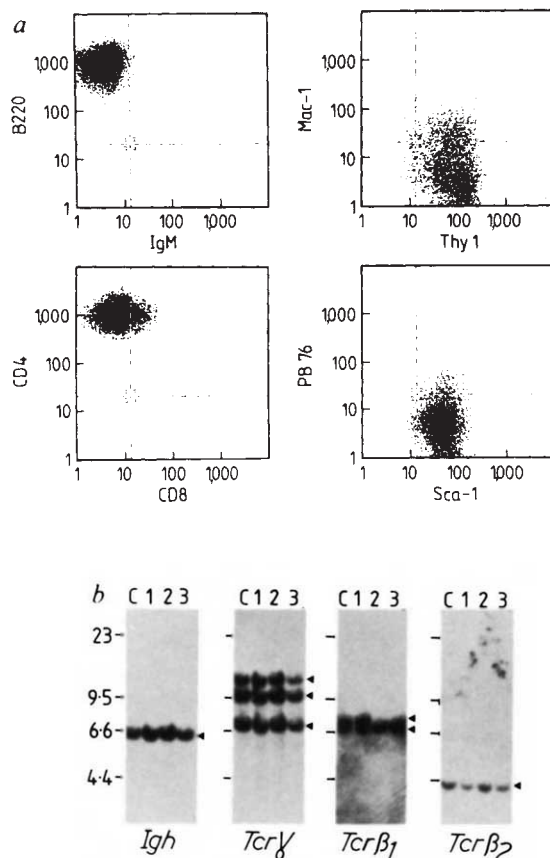


FIG. 3 Characterization of tumour cells in $E\mu$ -*bcl-2/myc* doubly transgenic mice. *a*, Surface phenotype analysis of cells from an enlarged mesenteric lymph node. *b*, Southern blot analysis of *Igh* and *Tcr* gene status in three independent tumours (1-3) and, as a control, liver (C) from a (C57BL/6 $\times$ SJL) F_1 mouse. Size markers in kilobases are shown to the left.

METHODS. Immunofluorescence staining and analysis have been described²³. The cell surface antigens and their identifying monoclonal antibodies were: CD45R (B220) (14.8 and RA3-6B2), IgM (5.1), Mac-1 (M1/70), Thy-1 (J1J and 30-H12), CD4 (GK1.5), CD8 (53-6.72), PB76 (G-5.2; ref. 22) and Sca-1 (E13-161.7; ref. 36). They were conjugated with fluorescein isothiocyanate or phycoerythrin or, alternatively, biotinylated and revealed with phycoerythrin-streptavidin. *EcoRI*-digested and size-fractionated tumour DNA was hybridized with J_{H4} and *Tcr* γ cDNA probes; *HindIII* digests were hybridized with β C1 and β C2 genomic probes (the germline β C1 fragments differ between C57BL/6 and SJL mice). The origin of the probes is detailed elsewhere³⁷; they were labelled with ³²P by random priming.

Received 4 June; accepted 13 September 1990.

1. Tsujimoto, Y., Finger, L. R., Yunis, J., Nowell, P. C. & Croce, C. M. *Science* **226**, 1097-1099 (1984).
2. Bakhshi, A. *et al.* *Cell* **41**, 889-906 (1985).
3. Cleary, M. L., Smith, S. D. & Sklar, J. *Cell* **47**, 19-28 (1986).
4. Fukuhara, S., Rowley, J. D., Variakojis, D. & Golomb, H. M. *Cancer Res.* **39**, 3119-3128 (1979).
5. Tsujimoto, Y., N. Ikegaki, N. & Croce, C. M. *Oncogene* **2**, 3-7 (1987).
6. Chen-Levy, Z., Nourse, J. & Cleary, M. L. *Molec. cell. Biol.* **9**, 701-710 (1989).
7. Reed, J. C., Tsujimoto, Y., Alpers, J. D., Croce, C. M. & Howell, P. C. *Science* **236**, 1295-1299 (1987).
8. Granger, W. B., Seto, M., Boutain, B., Goldman, P. & Korsmeyer, S. J. *J. clin. Invest.* **80**, 1512-1515 (1987).
9. Vaux, D. L., Cory, S. & Adams, J. M. *Nature* **335**, 440-442 (1988).
10. Nunez, G. *et al.* *J. Immun.* **144**, 3602-3610 (1990).
11. Tsujimoto, Y. *Oncogene* **4**, 1331-1336 (1989).
12. Reed, J. C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 3660-3664 (1990).
13. Reed, J. C., Haldar, S., Cuddy, M. P., Croce, C. M. & Makover, D. *Oncogene* **4**, 1123-1127 (1989).
14. Nunez, G. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 4589-4593 (1989).
15. McDonnell, T. J. *et al.* *Cell* **57**, 79-88 (1989).
16. Strasser, A. *et al.* *Curr. Topics Microbiol. Immun.* (in the press).
17. McDonnell, T. J. *et al.* *Molec. cell. Biol.* **10**, 1901-1907 (1990).
18. Cory, S. *Adv. Cancer Res.* **47**, 189-234 (1986).
19. Adams, J. M. *et al.* *Nature* **318**, 533-538 (1985).
20. Langdon, W. Y., Harris, A. W., Cory, S. & Adams, J. M. *Cell* **47**, 11-18 (1986).
21. Harris, A. W. *et al.* *J. exp. Med.* **167**, 353-371 (1988).
22. Schmidt, E. V., Pattengale, P. K., Weir, L. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6047-6051 (1988).
23. Strasser, A. *Eur. J. Immun.* **18**, 1803-1810 (1988).
24. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. *Science* **241**, 58-62 (1988).
25. Fredrickson, G. G. & Basch, R. S. *J. exp. Med.* **169**, 1473-1478 (1989).
26. Morse, H. C. III *et al.* *J. exp. Med.* **165**, 920-925 (1987).
27. Palacios, R. & Steinmetz, M. *Cell* **41**, 727-734 (1985).
28. Alessandrini, A., Pierce, J. H., Baltimore, D. & Desiderio, S. V. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1799-1803 (1987).
29. Sakaguchi, N. & Melchers, F. *Nature* **324**, 579-582 (1986).
30. Schatz, D. G., Oettinger, M. A. & Baltimore, D. *Cell* **59**, 1035-1048 (1989).
31. Tidmarsh, G. F., Heimfeld, S., Whitlock, C. A., Weissman, I. L. & Müller-Sieburg, C. E. *Molec. cell. Biol.* **9**, 2665-2671 (1989).
32. Langdon, W. L., Harris, A. W. & Cory, S. *Oncogene Res.* **3**, 271-279 (1988).
33. Pegoraro, L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **81**, 7166-7170 (1984).
34. Gauwerky, C. E., Haluska, F. G., Tsujimoto, Y., Nowell, P. C. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8548-8552 (1988).
35. Gauwerky, C. E., Huebner, K., Isobe, M., Nowell, P. C. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8867-8871 (1989).
36. Aihara, Y., Bühring, H.-J., Aihara, M. & Klein, J. *Eur. J. Immun.* **16**, 1391-1399 (1986).
37. Hariharan, I. K. *et al.* *Molec. cell. Biol.* **9**, 2798-2805 (1989).

ACKNOWLEDGEMENTS. We thank J. M. Adams for discussion and comments on the manuscript, D. Schatz and M. Oettinger for the *RAG-1* cDNA, F. Melchers for the λ 5 probe, K. Patane for animal husbandry, S. Mihajlovic for histological sections, M. Stanley for technical assistance and I. Collins for typing this manuscript. A.S. is the recipient of a fellowship from the Swiss National Science Foundation. This work was supported by the National Health and Medical Research Council of Australia and the US National Cancer Institute.

Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death

David Hockenbery, Gabriel Nuñez, Curt Millman, Robert D. Schreiber & Stanley J. Korsmeyer

Departments of Medicine, Molecular Microbiology, and Pathology, Howard Hughes Medical Institute, Washington University School of Medicine, St Louis, Missouri 63110, USA

THE t(14;18) chromosomal translocation of human follicular B-cell lymphoma juxtaposes the *bcl-2* gene with the immunoglobulin heavy chain locus¹⁻³. The *bcl-2* immunoglobulin fusion gene is markedly deregulated resulting in inappropriately elevated levels of *bcl-2* RNA and protein⁴⁻⁷. Transgenic mice bearing a *bcl-2* immunoglobulin minigene demonstrate a polyclonal expansion of resting yet responsive IgM-IgD B cells which display prolonged cell survival but no increase in cell cycling^{8,9}. Moreover, deregulated *bcl-2* extends the survival of certain haematopoietic cell lines following growth-factor deprivation^{10,11}. By using immunolocalization studies we now demonstrate that Bcl-2 is an integral inner mitochondrial membrane protein of relative molecular mass 25,000 (25k). Overexpression of Bcl-2 blocks the apoptotic death of a pro-B-lymphocyte cell line. Thus, Bcl-2 is unique among proto-oncogenes, being localized to mitochondria and interfering with programmed cell death independent of promoting cell division.

To establish the localization and biochemical role of Bcl-2 we generated a hamster monoclonal antibody, 6C8, monospecific for the 25K human Bcl-2 protein. Initial immunofluorescence studies using 6C8 indicated that Bcl-2 was located intracellularly rather than on the surface of B cells. When examined with a laser-scanning confocal microscope, the RL-7 cell line, which bears the t(14;18) translocation¹² was seen to have a punctate cytoplasmic distribution of Bcl-2-staining,

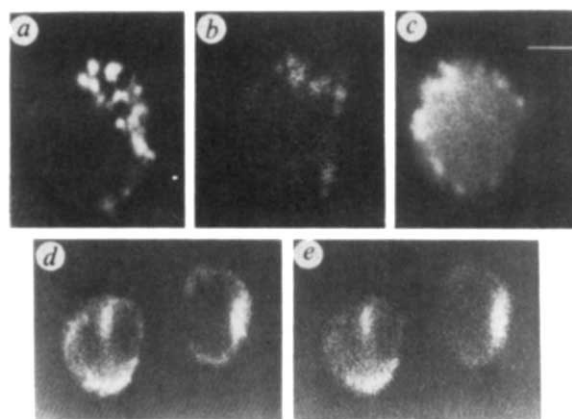
suggesting that the protein is localized to an organelle (Fig. 1a). The pattern of Bcl-2-staining resembled, that obtained using fluorescent probes such as streptavidin-Texas Red and rhodamine 123, which target mitochondria (Fig. 1b, c), but differed from that obtained using probes which target other organelles, such as acridine orange, which stains acidic lysosomal compartments. Dual fluorescence examination indicated that the staining with fluorescein isothiocyanate-labelled 6C8 is coincident with the distribution of mitochondria revealed by streptavidin-Texas red (Fig. 1d, e).

In earlier studies it has been concluded that Bcl-2 is localized to the inner surface of the plasma membrane or the perinuclear endoplasmic reticulum^{7,13}. Bcl-2 has a unique sequence with no substantial homology to other proto-oncogene products. It lacks an N-terminal hydrophobic signal sequence, or obvious internal signal motifs, that would target it to endoplasmic reticulum and subsequently the plasma membrane. However, Bcl-2 does possess a C-terminal hydrophobic 19-amino-acid stretch reminiscent of a membrane-spanning segment^{4,6,14}. Triton X-114 phase separation of B-cell lysates confirmed the lipophilic character of Bcl-2 and indicated that it is an integral membrane-spanning protein (ref. 7 and data not shown).

To determine which membrane compartment possessed Bcl-2 we subcellularly fractionated RL-7 cells. Bcl-2 was present in the heavy membrane fraction, which possesses most of the mitochondrial and lysosomal membranes, as assessed enzymatically (Fig. 2a). The light membrane fraction and the cytosolic supernatant did not contain substantial amounts of Bcl-2. Bcl-2 associated with the initial nuclear pellet was eliminated by sedimentation through 2 M sucrose to remove adherent organelle structures (Fig. 2a). Disrupted RL-7 cells were also separated on 1-2-M sucrose density gradients. Most Bcl-2 colocalized with mitochondria-rich fractions as opposed to the lighter fractions with plasma membrane or Golgi (Fig. 2b). Preparations of heavy membranes were separated on a 30% Percoll density gradient (Fig. 2c). The 25K Bcl-2 is found in fractions with mitochondrial succinate dehydrogenase activity, but not in fractions with the peak lysosomal β -hexosaminidase activity. Heavy membrane preparations were then loaded on a

FIG. 1 Fluorescence localization of Bcl-2 in RL-7 cells. a, Hamster anti-human Bcl-2 monoclonal antibody 6C8 and fluorescein-conjugated goat anti-hamster IgG. b, Streptavidin-Texas red conjugate. c, Rhodamine 123. d, e, Dual labelling with 6C8/fluorescein-conjugated goat anti-hamster IgG and streptavidin-Texas red, respectively. Size bar, 5 μ m.

METHODS. Monoclonal antibodies were produced in Armenian hamsters as described²⁰, using intact human Bcl-2 protein as immunogen. Bcl-2 was expressed in *Escherichia coli* strain JM103 using expression vector plasmid pMON 5743, a derivative of pMON 5515 (ref. 21). Bcl-2 cDNA No. 58 (ref. 6) was modified by polymerase chain reaction to introduce *Nco*I restriction sites, one 3' of the open reading frame and one including the ATG initiation codon, and subcloned into *Nco*I-digested pMON 5743. Protein expression was induced by nalidixic acid and Bcl-2 protein was eluted from SDS-gel slices before immunization. The 6C8 monoclonal antibody proved to be monospecific for human Bcl-2 and did not recognize murine Bcl-2 by western blotting⁹ or immunofluorescence. RL-7 cells were washed in PBS and adhered to AdhesioSlides (MM Developments, Ottawa, Canada) at 5×10^4 cells per spot. Cells were fixed in 4% paraformaldehyde in PBS for 20 min at room temperature, and permeabilized in 0.1% Triton X-100 in Tris-buffered saline, pH 7.4, with 0.1% BSA for 5 min. 6C8 hybridoma supernatant or control hamster monoclonal antibody was used with fluorescein-conjugated goat anti-hamster IgG (USB, Cleveland, Ohio) at a 1:1000 dilution. Antibodies were incubated for 20 min followed by three rinses in 0.1% Triton X-100 in PBS. Streptavidin-Texas Red conjugate (BRL, Gaithersburg, Maryland) was incubated with fixed, permeabilized cells at a 1:250 dilution for 1 h and rinsed as above. Rhodamine 123 at a concentration of $10 \mu\text{g ml}^{-1}$ in DMEM medium (Molecular Probes, Eugene, Oregon) was incubated for 30 min with freshly adhered cells at 37 °C and rinsed three times with DMEM before fixing. Streptavidin conjugates bind to biotinylated carboxylase enzymes²² present in mitochondria. Photographs were taken using a BioRad Lasersharp MRC500 scanning confocal microscopy system with filters for dual fluorescence and Texas Red emissions. A Zeiss Axioplan microscope was used for fluorescent microscopy.



9–35% Nycodenz gradient. The amount of Bcl-2 corresponded with the mitochondrial succinate dehydrogenase activity, whereas there was no Bcl-2 in the peak fraction of endoplasmic reticulum (Fig. 2d). Finally, the mitochondrial fraction of a sucrose step gradient was disrupted into fractions of inner and outer mitochondrial membrane by hypotonic lysis. Immunoblots indicated that Bcl-2 was present exclusively in the pellet of the inner mitochondrial mitoplast with all of the succinate dehydrogenase activity. In contrast, the supernatant with all of the outer mitochondrial membrane monoamine oxidase activity lacked Bcl-2 (Fig. 2e).

Deregulated Bcl-2 expression extends the survival of the interleukin-3 (IL-3)-dependent pro-B lymphocyte line, FL5.12, when

deprived of growth factor¹¹. The species-specific anti-human Bcl-2 monoclonal antibody allowed us to examine the targeting of transfected human Bcl-2 in murine FL5.12 (Fig. 3a). A vector based on the long terminal repeat of the splenic focus-forming virus (SFFV)¹⁵ overproduced the normal 25K human Bcl-2 protein, which demonstrated the same cytoplasmic immunofluorescence pattern as endogenous Bcl-2 (not shown). Subcellular fractionation of FL5.12-bearing this vector (SFFV-Bcl-2 nl) indicated that the introduced human Bcl-2 correctly targeted to the heavy membrane fraction (Fig. 3a) and localized to the mitochondrial fractions of a sucrose density gradient (not shown). Given the mitochondrial localization of Bcl-2, it was important to examine the mechanism of death of FL5.12 cells.

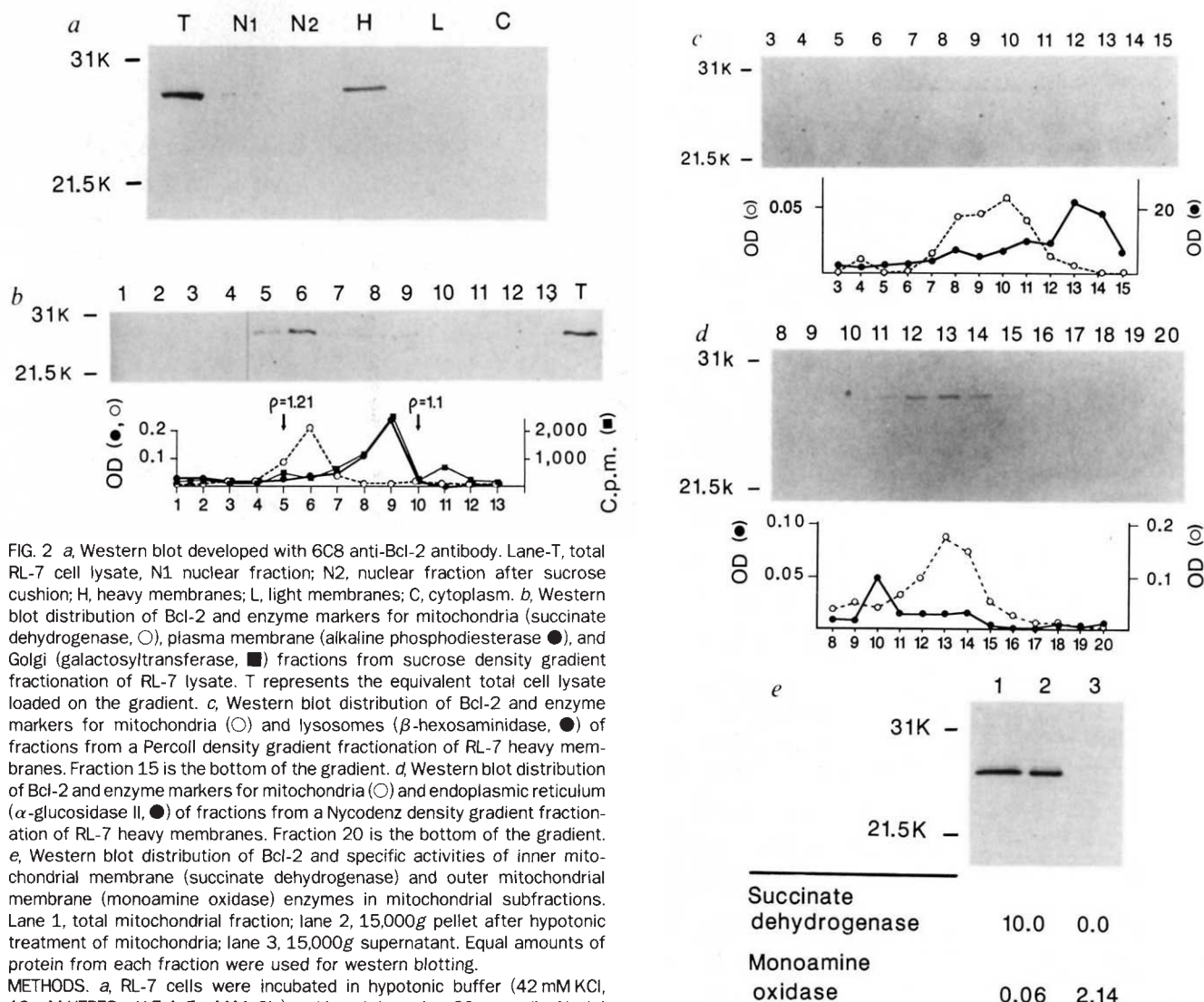


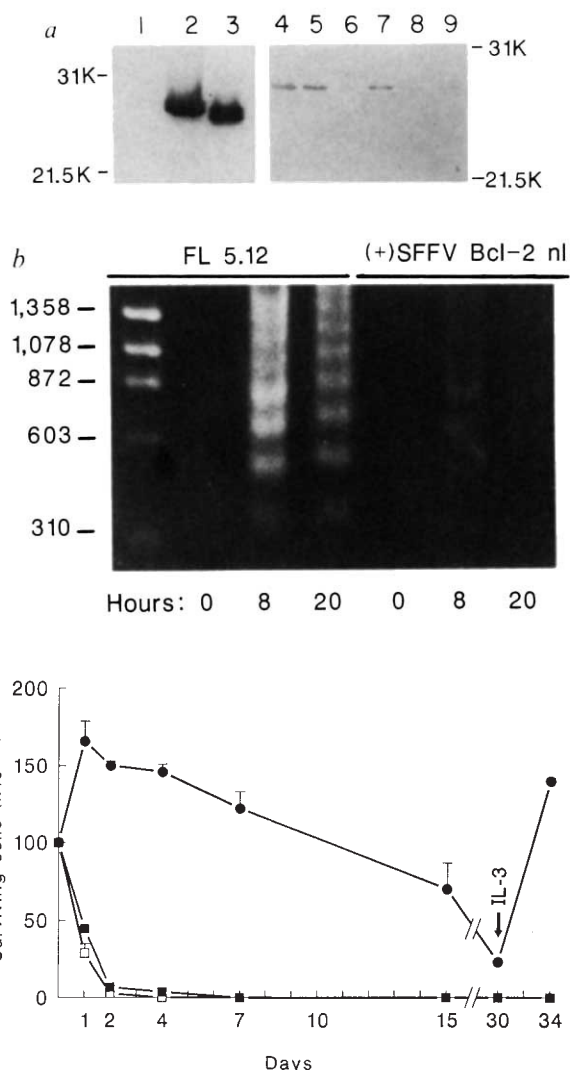
FIG. 2 a, Western blot developed with 6C8 anti-Bcl-2 antibody. Lane-T, total RL-7 cell lysate; N1 nuclear fraction; N2, nuclear fraction after sucrose cushion; H, heavy membranes; L, light membranes; C, cytoplasm. b, Western blot distribution of Bcl-2 and enzyme markers for mitochondria (succinate dehydrogenase, ○), plasma membrane (alkaline phosphodiesterase, ●), and Golgi (galactosyltransferase, ■) fractions from sucrose density gradient fractionation of RL-7 lysate. T represents the equivalent total cell lysate loaded on the gradient. c, Western blot distribution of Bcl-2 and enzyme markers for mitochondria (○) and lysosomes (β -hexosaminidase, ●) of fractions from a Percoll density gradient fractionation of RL-7 heavy membranes. Fraction 15 is the bottom of the gradient. d, Western blot distribution of Bcl-2 and enzyme markers for mitochondria (○) and endoplasmic reticulum (α -glucosidase II, ●) of fractions from a Nycodenz density gradient fractionation of RL-7 heavy membranes. Fraction 20 is the bottom of the gradient. e, Western blot distribution of Bcl-2 and specific activities of inner mitochondrial membrane (succinate dehydrogenase) and outer mitochondrial membrane (monoamine oxidase) enzymes in mitochondrial subfractions. Lane 1, total mitochondrial fraction; lane 2, 15,000g pellet after hypotonic treatment of mitochondrial fraction; lane 3, 15,000g supernatant. Equal amounts of protein from each fraction were used for western blotting.

METHODS. a, RL-7 cells were incubated in hypotonic buffer (42 mM KCl, 10 mM HEPES, pH 7.4, 5 mM $MgCl_2$) and lysed through a 30-g needle. Nuclei were pelleted at 200g for 10 min. The supernatant was centrifuged at 10,000g for 10 min to collect the heavy membrane pellet. That supernatant was centrifuged at 150,000g for 90 min to yield the light membrane pellet and final cytoplasmic supernatant. An aliquot of the nuclear pellet was overlaid on a 2.0M sucrose cushion. The N2 nuclear pellet was collected after centrifuging at 150,000g for 90 min. Western blotting was as previously described⁹. b, RL-7 lysate was loaded on a linear 1–2M sucrose gradient. The gradient was centrifuged at 25,000 r.p.m. (110,000g maximum) for 3 h at 4 °C. Succinate dehydrogenase was assayed as described²³. Alkaline phosphodiesterase was assayed in 20 mM CAPS (3-cyclohexylamino-1-propane-sulphonic acid), pH 10.6, 0.04% Triton X-100 with 1 mM thymidine 5' monophosphate-*p*-nitrophenyl ester as substrate. Galactosyltransferase was determined as described²⁴. Yields from the gradient for each enzyme were: succinate dehydrogenase, 109% alkaline phosphodiesterase, 136% and galactosyltransferase, 132%. c, The heavy mem-

brane fraction was mixed with 30% Percoll in 0.25 M sucrose. A gradient was formed *in situ* by centrifuging at 48,000g for 60 min at 4 °C. β -hexosaminidase was assayed as described²⁴. d, A heavy membrane fraction layered over a 9–35% linear Nycodenz gradient in 0.25 M sucrose. The gradient was centrifuged at 28,000 r.p.m. (140,000g maximum) for 2 h in an SW28 rotor. α -glucosidase II was assayed as described²⁵. e, The mitochondrial fraction at the 1.15–1.20 g ml⁻¹ interface from a sucrose step gradient was centrifuged at 15,000g for 10 min. The pellet was resuspended in 1 ml 20 mM KPO_4 , pH 7.4, 0.02% BSA and incubated at 4 °C for 20 min²⁶. ATP and $MgCl_2$ were added to final concentrations of 1 mM. The suspension was centrifuged at 15,000g for 10 min. Monoamine oxidase was assayed as described²⁷. Enzymatic assays of the 15,000g pellet indicated that it contained 107% of total succinate dehydrogenase, <1% of total monoamine oxidase, 5% of total β -hexosaminidase, 3.4% of total galactosyltransferase, and 3.5% of total α -glucosidase II.

FIG. 3 *a*, Left panel: species specificity of 6C8 monoclonal antibody. Western blot of cell lysates in: lane 1, FL5.12; lane 2, FL5.12 stably transfected with SFFV Bcl-2 nl possessing intact human Bcl-2; lane 3, FL5.12 stably transfected with SFFV Bcl-2 Δ C-22 lacking the C-terminal 22 amino acids. Right panel: western blot of subcellular fractions of FL5.12 bearing the SFFV Bcl-2 nl vector developed with 6C8 anti-Bcl-2 antibody. Lane 4, total cell lysate; 5, nuclear fraction; 6, nuclear fraction after sucrose cushion; 7, heavy membranes; 8, light membranes; 9, cytoplasm. *b*, Agarose gel electrophoresis of oligonucleosomal-length DNA fragments from IL-3-deprived FL5.12 cells or cells transfected with SFFV Bcl-2 nl. *c*, Survival in days after IL-3 deprivation of FL5.12 cells transfected with the control construct SFFV which contained the neomycin resistance gene NEO ($\square$), SFFV Bcl-2 nl ($\bullet$) or construct containing the *Ttg-1* gene ($\blacksquare$)²⁸. Data is expressed as means of triplicate cultures $\pm$ s.d.

METHODS. Complementary DNA no. 58 (ref. 6) containing entire human *bcl-2* opening reading frame was inserted into the *EcoRI* cloning site of SFFV Neo vector¹⁵. A truncated *bcl-2* sequence, Bcl-2 Δ C-22, was obtained by polymerase chain reaction, using cDNA no. 58 and modified primer which introduced a stop codon at codon 218 (GAGGAATTCTCAGAGACAGCCAGAGAA), subsequently confirmed by dideoxy-chain-termination sequencing. Similarly a 1.2-kilobase cDNA containing the opening reading frame of the *ttg-1* putative oncogene was inserted into the *EcoRI* cloning site of SFFV Neo. Constructs were transfected into FL5.12 cells by electroporation (200 V, 900 mF). Cells were selected for neomycin resistance in 1 mg ml⁻¹ G418. FL5.12 or FL5.12 SFFV Bcl-2 nl cells (4×10^6) at 0, 8 and 20 hours after IL-3 deprivation were lysed in 0.5% Triton X-100, 5 mM Tris buffer pH 7.4, 20 mM EDTA for 20 min at 4 °C. After centrifugation at 27,000g for 15 min supernatants were extracted with phenol-chloroform and precipitated in ethanol. One half of each sample was electrophoresed on a 1.2% agarose gel. RNase A (20 μ g ml⁻¹) was incubated with the gel at 37 °C for 3 h before staining with ethidium bromide.



Serial examination of FL5.12 cells after they had been deprived of IL3 revealed that death was by apoptosis^{16,17}, in which plasma membrane blebbing and volume loss was followed by nuclear condensation and an endonucleolytic cleavage of DNA into fragments of oligo-nucleosomal length (Fig. 3*b*). Moreover, the death of FL5.12 cells can be delayed by concentrations of cycloheximide that block new protein synthesis, arguing that this is an active death programme (not shown). The presence of deregulated Bcl-2 blocks this death programme, including the endonucleolytic cleavage of DNA (Fig. 3*b*). FL5.12 cells bearing the SFFV-Bcl-2 nl vector remained alive but in the G0 stage of the cell cycle for more than 30 days after IL-3 deprivation, and re-entered the cell cycle after the addition of the growth factor (Fig. 3*c*).

The localization of Bcl-2 to the inner mitochondrial membrane is novel for proteins with an oncogenic role. It is not known how Bcl-2 interferes with the death by apoptosis of FL5.12 cells deprived of IL-3. Haldar *et al.*¹⁸ have proposed that Bcl-2 is a G protein. But Bcl-2 does not possess highly conserved nucleotide-binding motifs, and recent studies by Monica *et al.*¹⁹ indicate that Bcl-2 does not bind GTP. Indeed the native protein recognized by our 6C8 monospecific monoclonal antibody does not bind GTP. The localization of Bcl-2 suggests that the main metabolic functions of the inner mitochondrial membrane, which include oxidative phosphorylation, electron and metabolite transport, are involved in the survival mechanism. These studies suggest a prominent role for Bcl-2 in a B-cell survival pathway. $\square$

Received 14 June; accepted 20 September 1990.

1. Tsujimoto, Y., Finger, L. R., Yunis, J., Nowell, P. C. & Croce, C. M. *Science* **226**, 1097-1099 (1984).
2. Bakshi, A. *et al. Cell* **41**, 889-906 (1985).
3. Cleary, M. L. & Sklar, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7439-7443 (1985).
4. Cleary, M. L., Smith, S. D. & Sklar, J. *Cell* **47**, 19-28 (1986).
5. Graninger, W. B., Seto, M., Boutain, B., Goldman, P. & Korsmeyer, S. J. *J. clin. Invest.* **80**, 1512-1515 (1987).
6. Seto, M. *et al. EMBO J.* **7**, 123-131 (1988).
7. Chen-Levy, Z., Nourse, J. & Cleary, M. L. *Molec. cell. Biol.* **9**, 701-710 (1989).
8. McDonnell, T. J. *et al. Cell* **57**, 79-88 (1989).
9. McDonnell, T. J. *et al. Molec. cell. Biol.* **10**, 1901-1907 (1990).
10. Vaux, D. L., Cory, S. & Adams, J. M. *Nature* **335**, 440-442 (1988).
11. Nunez, G. N. *et al. J. Immun.* **144**, 3602-3610 (1990).
12. Lipford, E. *et al. Blood* **70**, 1816-1823 (1987).
13. Tsujimoto, Y., Ikegaki, N. & Croce, C. M. *Oncogene* **2**, 3-7 (1987).
14. Tsujimoto, Y. & Croce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5214-5218 (1986).
15. Fuhbrigge, R. C., Fine, S. M., Unanue, E. R. & Chaplin, D. D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5649-5653 (1988).
16. Wyllie, A. H., Morris, R. G., Smith, A. L. & Dunlop, D. J. *Path.* **142**, 67-77 (1984).
17. Williams, G. T., Smith, C. A., Spooner, E., Dexter, T. M. & Taylor, D. R. *Nature* **343**, 76-79 (1990).
18. Haldar, S., Beatty, C., Tsujimoto, Y. & Croce, C. M. *Nature* **342**, 195-198 (1989).
19. Monica, K., Chen-Levy, Z. & Cleary, M. L. *Nature* **346**, 189-191 (1990).
20. Schrieber, R. D., Hicks, L. J., Celada, A., Buchmeier, N. A. & Gray, P. W. *J. Immun.* **134**, 1609-1618 (1985).
21. Olins, P. O., Devine, C. S., Rangwala, S. H. & Karka, K. S. *Gene* **73**, 227-235 (1988).
22. Robinson, B. H. *et al. Am. J. hum. Genet.* **40**, 50-59 (1987).
23. Pennington, R. J. *Biochem. J.* **80**, 649-654 (1961).
24. Messner, D. J., Griffiths, G. & Kornfeld, S. *J. cell. Biol.* **108**, 2149-2162 (1989).
25. Micheal, J. M. & Kornfeld, S. *Archiv Biochem. Biophys.* **199**, 249-258 (1980).
26. Murthy, M. S. R. & Pande, S. V. *Proc. natn. Acad. Sci. U.S.A.* **84**, 378-382 (1987).
27. Schraitman, C., Erwin, V. G. & Greenwalt, J. W. *J. cell. Biol.* **32**, 719-735 (1967).
28. McGuire, E. A. *et al. Molec. cell. Biol.* **9**, 2124-2132 (1989).

ACKNOWLEDGEMENTS. We thank the Lucille Markey Image Analysis Facility for laser scanning confocal microscopy, D. Schafer for streptavidin staining protocol, P. Olins for the pMON 5743 vector, and A. Schwartz, K. Sheehan, D. Messner and S. Kornfeld for helpful discussions. This research was supported by the American Cancer Society (D.H.) and by NIH.

Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant

Claude W. dePamphilis* & Jeffrey D. Palmer

Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

PHOTOSYNTHESIS is the hallmark of plant life and is the only plastid metabolic process known to be controlled by plastid genes^{1,2}. The complete loss of photosynthetic ability, however, has occurred on several independent occasions in parasitic flowering plants^{3,4}. Some of these plants are known to lack chlorophyll and certain photosynthetic enzymes⁴, but it is not known to what extent changes have occurred in the genes encoding the photosynthetic apparatus or whether the plants even maintain a plastid genome. Here we report that the nonphotosynthetic root parasite *Epifagus virginiana* has a plastid chromosome only 71 kilobases in size, far smaller than any previously characterized land plant plastid genome⁵. The *Epifagus* plastid genome has lost most, if not all, of the 30 or more chloroplast genes for photosynthesis and most of a large family of plastid genes, the *ndh* genes, whose products may be involved in a plastid respiratory chain⁶⁻⁹. The extensive changes in *Epifagus* plastid gene content must have occurred in a relatively short time (5–50 × 10⁶ yr), because *Striga asiatica*, a related photosynthetic parasite, has a typical complement of chloroplast genes for photosynthesis and chlororespiration. The plastid genome of *Epifagus* has retained transcribed ribosomal RNA and ribosomal protein genes, suggesting that it expresses one or more gene products for plastid functions not related to photosynthesis.

Epifagus virginiana (beechdrops) is a member of the *Orobanchaceae*, a family of root-parasitic angiosperms believed to lack chlorophyll and photosynthetic function^{4,10}, and thus termed 'holoparasitic'. Like most other *Orobanchaceae* that have been examined¹¹⁻¹³, *Epifagus* lacks detectable levels of chlorophyll and the photosynthetic enzyme ribulose biphosphate carboxylase/oxygenase, and acquires its carbon entirely from a photosynthetic host (the beech tree, *Fagus grandifolia*) (ref. 14 and C.W.D., J. Seeman, N. Bowlby, J. A. Teeri and J.D.P., unpublished data). Despite this lack of photosynthetic ability, *Epifagus* contains plastids, some of which accumulate starch¹⁵.

Eighty-three cloned fragments representing the entire plastid genome of *Nicotiana tabacum* were used in filter blot hybridizations to construct a physical and gene map of the *Epifagus* plastid DNA (ptDNA). The genome is a 71-kilobase (kb) circular chromosome with a ribosomal-RNA-encoding inverted repeat (IR) structure characteristic of most ptDNAs¹⁶ and with many other plastid genes colinear with those of *Nicotiana* (Fig. 1). The *Epifagus* genome is the smallest by far of over 1,000 now-examined land plant ptDNAs (range 120–220 kb; ref. 5), and, moreover, has only about one-third the sequence complexity of a typical plastid genome (46 kb versus 130 kb for *Nicotiana* ptDNA; ref. 2).

To illustrate several of the conserved and deleted sequences in *Epifagus* ptDNA, and to examine the distribution of photosynthetic and *ndh* genes in plants related to *Epifagus*, we hybridized various gene probes from *Nicotiana* to DNAs of *Epifagus*, *Nicotiana*, *Antirrhinum majus* and *Striga asiatica*. The latter two species are, respectively, nonparasitic and hemiparasitic (that is, a photosynthetic parasite) members of the Scrophulariaceae family, from which the *Orobanchaceae* was derived³. Each of the gene probes hybridized strongly to the photosynthetic plants,

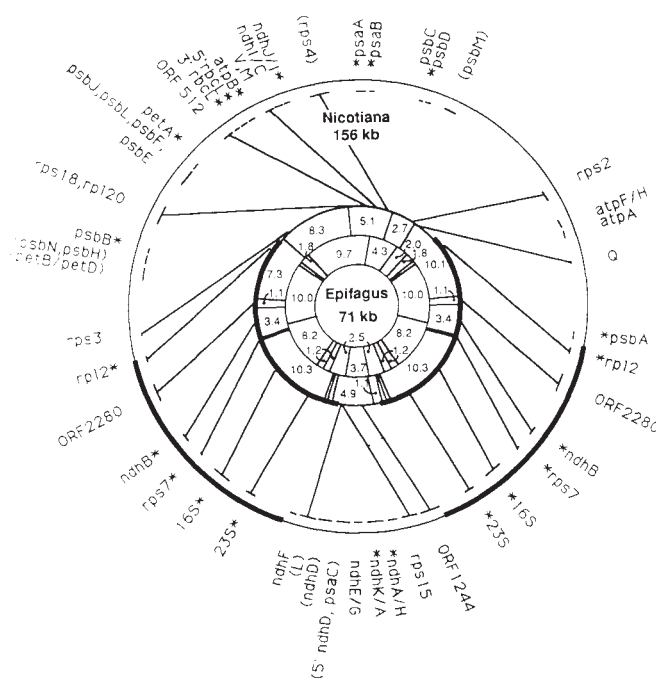


FIG. 1 The 71-kb ptDNA of *Epifagus virginiana* has lost numerous photosynthetic genes and *ndh* genes while retaining rRNA and tRNA genes, ribosomal protein genes, ORFs and an IR. The two inner circles show restriction site maps (*Clal*, outer; *HindIII*, inner), for *Epifagus* ptDNA; outer circle a partial gene map for tobacco ptDNA². Gene nomenclature follows ref. 5. IR segments in both genomes are indicated with thickened lines. Gene probes from *Nicotiana* that hybridized to *Epifagus* ptDNA are shown by lines connecting the two maps; other indicated genes from *Nicotiana* failed to hybridize to *Epifagus* ptDNA. All probes are gene-specific except those given in parentheses. Hybridization results illustrated in Fig. 2 are indicated by an asterisk. Transfer RNA genes are indicated by the single-letter code for their cognate amino acids.

METHODS. Total DNA was extracted from above ground tissue of *Epifagus* (Washtenaw) as described²⁷ and digested with *XhoI*, *Clal*, *HindIII*, *SacI*, *SmaI*, *PstI* and *Sall* singly and in double digest with *XhoI*. The digested DNAs were separated on 0.8% agarose gels and transferred to nylon filters (Zetabind; AMF Cuno) for initial hybridizations with 60 contiguous subclones of the entire ptDNA of *Nicotiana*. Restriction enzyme digestion, agarose gel electrophoresis, Southern transfer (double-sided dry blots), preparation of ³²P-labelled probes and hybridization were as described²⁸. Hybridizations were at 65 °C and after hybridization filters were washed once at room temperature and three times at 65 °C in 2 × SSC, 0.5% SDS. Lower stringency hybridizations and washes at 50 °C and 4 × SSC for the genes *rbcl* and *psbA* also failed to detect a signal from *Epifagus*. Construction of the map from the hybridization results was as described²⁹. Complete restriction maps for each of seven enzymes showing size and location of each deletion and conserved region will be presented elsewhere. A list of the probes used for the restriction site mapping may be obtained from the authors. After construction of the restriction site map, hybridization results from small probes specific for one or a few genes, were used to identify gene locations on the *Epifagus* map. The probes used for gene mapping are listed below, followed by their coordinates from the *Nicotiana* plastid genome³⁰: *psbA*, 419–1731; (*Q*), 6,652–7,637; *atpA*, 10,611–12,060; *atpF/H*, 12,305–13,913; (*rps2*), 16,000–17,446; (*psbM*), 29,820–31,918; *psbD*, 34,656–35,492; *psbC*, 35,839–36,455; *psaA*, 41,611–43,426; (*rps4*), 45,076–48,602; *ndhJ/I*, 51,205–52,330; (*ndhI/C*), 52,330–53,599; *V/M*, 53,599–54,870; *atpB*, 55,921–56,836; 5' *rbcl*, 56,836–58,047; 3' *rbcl*, 58,047–59,305; ORF 512, 59,305–60,654; *petA*, 64,101–65,302; *psbJ*, *psbL*, *psbE*, 65,302–67,129; (*rps18*, *rpl20*), 70,773–74,167; *psbB*, 75,822–76,545; (*psbN*, *psbH*), 76,545–77,375; *petB/petD*, 77,375–79,188; *rps3*, 85,250–85,632; *rpl2*, 86,681–87,725; ORF 2280, 91,923–94,562; *ndhB*, 96,981–98,527; *rps7*, 99,726–99,987; 16S RNA, 101,532–104,801; 23S RNA, 107,136–108,816; *ndhF*, 111,924–113,113; (*L*), 113,119–116,171; (*ndhD*), 116,171–118,604; *psaC*, 118,604–119,520; *ndhE/G*, 119,520–120,809; *ndhK/A*, 120,809–123,676; *ndhA/H*, 123,676–124,907; (*rps15*), 124,907–127,440; ORF 1244, 127,440–130,600. For *psaB*, we used a 1.6-kb *BamHI* gene-internal fragment from the ptDNA of *Spinacia oleracea*.

* Present address: Department of General Biology, Vanderbilt University, Nashville, Tennessee 37240, USA.

including the hemiparasite *Striga* (Fig. 2). In addition, most of the gene probes derived from the *Nicotiana* IR hybridized strongly to the IR region of *Epifagus* ptDNA. Ribosomal protein genes and various open reading frames (ORFs) hybridized less strongly to the *Epifagus* single-copy regions. By contrast, only one of the 22 photosynthetic genes examined, *psaA*, was detected in *Epifagus* DNA. Because the *psaA* hybridization signal did not map to any location on the *Epifagus* plastid genome (Figs 1, 2), and because its strength appears too great for a single-copy nuclear gene, we suspect that this represents a sequence that has been incorporated into the mitochondrial genome^{17,18}.

The 10 or more *ndh*-like genes^{7,8,19,20} represent the largest family of plastid genes without a clearly defined function. These sequences have significant amino-acid similarity to mitochondrial and nuclear genes encoding components of the mitochondrial respiratory chain NADH dehydrogenase. The products of these genes could be involved in a membrane-bound respiratory

process present in the chloroplast (chlororespiration)^{7,20}, but the only direct evidence for this comes from the alga *Chlamydomonas reinhardtii*^{6,9}. That these genes are conserved across a variety of land plants² and are expressed⁷ supports the view that such a chloroplast respiratory chain also exists in land plants. All *ndh* probes from the tobacco small and large single-copy regions failed to hybridize to *Epifagus* ptDNA; only weak hybridization was detected with the IR gene *ndhB* (Fig. 2). Because mitochondrial respiration is one of the physiological functions that is expected to remain intact in parasitic plants⁴, the loss of the *ndh* genes from the *Epifagus* plastid genome makes it unlikely that products of these genes are essential components of a mitochondrial NADH-dehydrogenase complex. Instead, their loss is consistent with their function being tied to photosynthetic metabolism and, hence, to the chloroplast of green plants.

Transcripts from *Epifagus* of plastid 16S rRNA, 23S rRNA and ribosomal protein mRNA were detected on northern blots (Fig. 3 and data not shown). As a proportion of total cellular RNA, *Epifagus* fruits contain about one-fiftieth as much plastid rRNA as do *Nicotiana* leaves, whereas mitochondrial rRNAs are present in nearly equal levels in the two tissues (Fig. 3). As expected, given the evidence from the Southern blots (Figs 1, 2), transcripts of *rbcL* and *psbA* were not detected in *Epifagus*, whereas abundant transcripts of these genes were observed in *Nicotiana*, *Antirrhinum* and *Striga* (data not shown).

We suggest that a rapid loss of photosynthetic and chloro-respiratory genes followed the relaxation of selection for photosynthetic capacity in ancestors of *Epifagus*, irreversibly committing the plant to a heterotrophic existence. The range of times since the holoparasite *Epifagus* and the hemiparasite *Striga* shared a common ancestor probably does not extend beyond the limits of 5×10^6 years ago (oldest fossil record of Orobanchaceae) and 50×10^6 years ago (oldest fossil record of Scrophulariales; ref. 21). At either extreme, this implies much more rapid change in gene content than is normally found for ptDNA^{2,16,22}. This evolutionary dynamic is also in stark contrast to that of the mitochondrial and nuclear genomes of plants, which are much more prone to the accumulation of foreign sequences, including a variety of chloroplast-derived sequences, and the retention of nonfunctional endogenous sequences^{17,18}. The *psaA* sequence of *Epifagus*, which does not presently reside in the plastid genome, may be an example of one such nonfunctional gene that has gained a longer lifespan outside the plastid genome than within it.

An interesting contrast to the photosynthetic gene losses seen in *Epifagus* occurs in *Astasia longa*, a nonphotosynthetic alga morphologically similar to *Euglena gracilis*. A circular 73-kb DNA isolated from *Astasia* has major structural features similar to that of *Euglena* ptDNA, but lacks several chloroplast photosynthetic genes²³. Unlike *Epifagus*, *Astasia* retains an intact and expressed *rbcL* gene²⁴. It is possible that the *Epifagus* plastid genome has gone further in its reduction of unnecessary information. More likely, the *rbcL* gene product could have a function in *Astasia* that is either not found in *Epifagus* or supplied by some other gene product. The presence of transcribed copies of

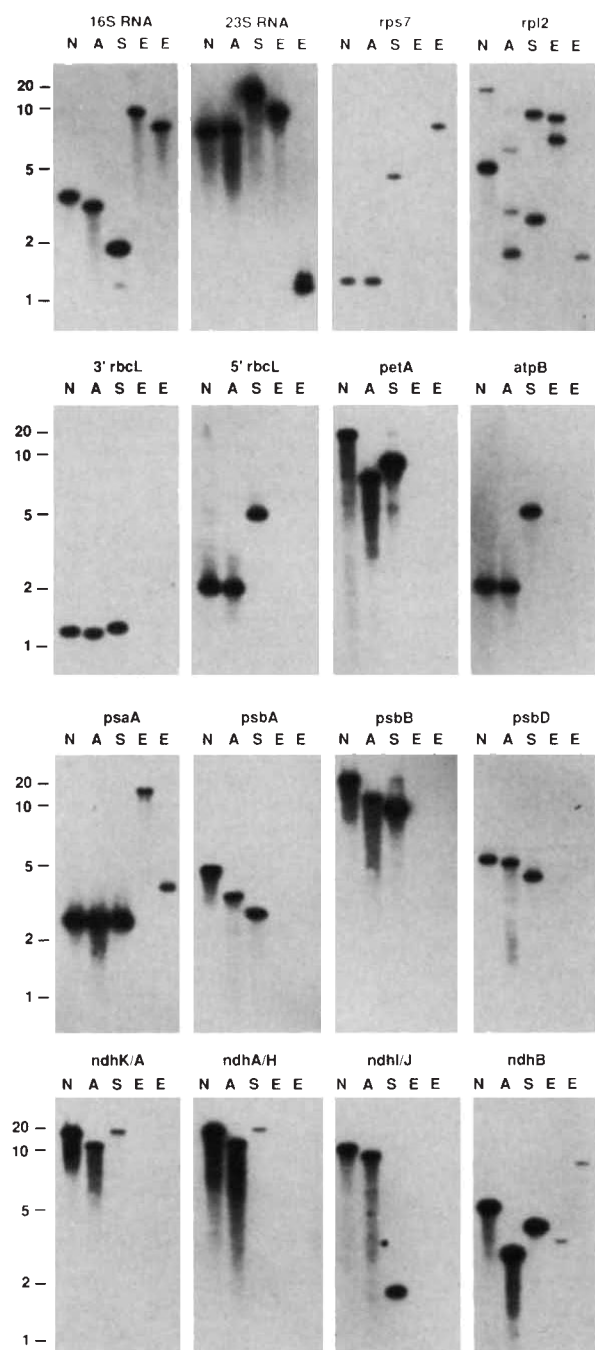


FIG. 2 Southern blots illustrating presence of chloroplast ribosomal RNA and protein genes and absence of numerous chloroplast photosynthetic and *ndh* genes from *Epifagus*. Each of 16 panels shown is a Southern blot of ptDNA of *Nicotiana tabacum* (N), *Antirrhinum majus* (A), and *Striga asiatica* (S), all cut with *Bam*HI, and two lanes (E) of *Epifagus virginiana* total DNA cut with *Clal* (left) and *Hind*III (right), and hybridized to the indicated gene-specific probe (see legend to Fig. 1 for probe coordinates). Fragment sizes are marked at left in kb. A 300-base-pair (bp) *Clal* fragment hybridizing to *rps7* cannot be seen here.

METHODS. As in Fig. 1; in addition, ptDNAs were extracted from sucrose gradient-purified chloroplasts as described²⁹. Probes for hybridization were isolated from plasmid clones in 1.0% low-gelling-temperature agarose and nick-translated directly in agarose gels or following separation from gel with GeneClean (Bio 101, La Jolla, California).

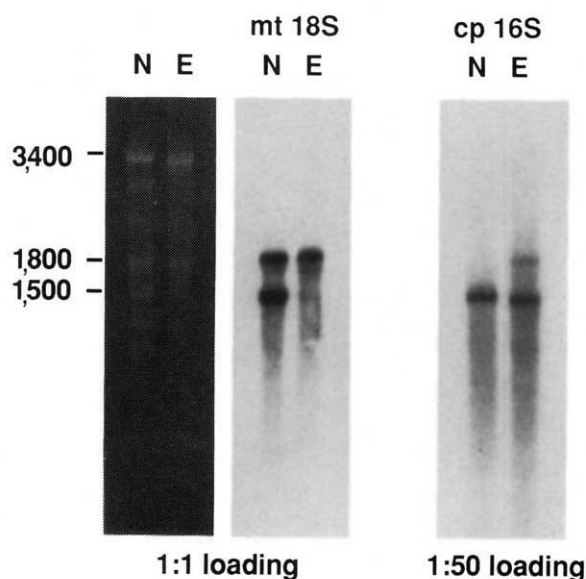


FIG. 3 Northern blot analysis illustrating presence of mitochondrial (mt) and chloroplast (cp) transcripts in *Epifagus*. Total RNA from leaves of *Nicotiana tabacum* (N) and fruits of *Epifagus virginiana* (E) were fractionated on a 1.5% agarose-formaldehyde gel²⁸. For 1:1 loadings we adjusted the *Nicotiana* and *Epifagus* RNAs so that the cytoplasmic rRNA bands were of about equal intensity. *Nicotiana* RNA was diluted 50-fold in water for the 1:50 loading. RNAs were transferred to nylon filters and hybridized with clones containing the mitochondrial 18S rRNA gene from *Brassica campestris* and the chloroplast 16S rRNA gene from *Nicotiana*. Sizes of RNA bands in nucleotides are indicated for the 28S and 18S cytoplasmic rRNA and the chloroplast 16S rRNA (1,500 bp). The smaller band in the *Nicotiana* 18S hybridization and the larger band in the *Epifagus* 16S hybridization represent cross-hybridization to the 16S and 18S RNAs, respectively, and reflect the well-established sequence similarities between plant mitochondrial 18S rRNA and chloroplast 16S rRNA³².

METHODS. Total RNA was extracted from whole green leaves of *Nicotiana* and fruits of *Epifagus*, frozen and powdered in liquid nitrogen, then homogenized in guanidinium isothiocyanate buffer and pelleted through RNase-free 5.66 M CsCl (ref. 31). Prehybridization and hybridization (65 °C) were in 0.5% SDS, 2 × SSC, 5% dextran sulphate, 100 µg ml⁻¹ carrier DNA²⁸. Preparation of ³²P-labelled gel-isolated probes and post-hybridization washes were as in Figs 1 and 2.

rbcL and *rbsS* in the holoparasite *Cuscuta*²⁵ might also be explained in either of these ways.

Why does *Epifagus* maintain a plastid genome? The highly selective retention of components of the translational apparatus in the *Epifagus* plastid genome and its transcription suggest that the genome is retained to produce one or more gene products involved in nonphotosynthetic plastid metabolic processes^{1,26}. Although these genes may include one or more of the plastid-encoded ORFs with so-far unknown function², the *ndh* genes are clearly not among them. □

Received 1 July; accepted 1 October 1990.

- Jensen, R. G. in *The Biochemistry of Plants* Vol. 1 (eds Stumpf, P. K. & Conn, E. E.) 273–313 (Academic, New York, 1980).
- Sugiura, M. A. *Rev. Cell Biol.* **5**, 51–70 (1989).
- Kuijt, J. *The Biology of Parasitic Flowering Plants* (University of California Press, 1969).
- Press, M. C. *Trends ecol. Evol.* **4**, 258–263 (1989).
- Palmer, J. D. in *Plant Gene Research*, Vol. 6: *Organelles* (1991).
- Bennoun, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4352–4356 (1982).
- Matsubayashi, T. et al. *Molec. gen. Genet.* **210**, 385–393 (1987).
- Fearnley, I. M., Runswick, M. J. & Walker, J. E. *EMBO J.* **8**, 665–672 (1989).
- Maione, T. E. & Gibbs, M. *Pl. Physiol.* **80**, 364–368 (1986).
- Musselman, L. J. A. *Rev. Phytopathol.* **18**, 463–489 (1980).
- de la Harpe, A. C., Visser, J. H. & Grobbelaar, N. Z. *Pflanzenphysiol.* **103**, 265–275 (1981).
- Press, M. C., Shah, N. & Stewart, G. R. in *Proceedings of a Workshop on Biology and Control of Orobanche* (ed. ter Borg, S. J.) 96–106 (LHV/VPO, Wageningen, The Netherlands, 1986).
- Bricaud, C. H., Thalouarn, P. & Renaudin, S. *J. Pl. Physiol.* **125**, 367–370 (1986).
- Musser, R. L. thesis (Duke University, North Carolina, 1979).
- Walsh, M. A., Rechel, E. A. & Popovich, T. M. *Am. J. Bot.* **67**, 833–837 (1980).
- Palmer, J. D. A. *Rev. Genet.* **19**, 325–354 (1985).
- Schuster, W. & Brennick, A. *Pl. Sci.* **54**, 1–10 (1988).
- Lonsdale, D. M. *Biochem. Plants* **15**, 229–295 (1989).

- Nixon, P. J. et al. *J. biol. Chem.* **264**, 14129–14135 (1989).
- Wu, M., Nie, Z. Q. & Yang, J. *Pl. Cell* **1**, 551–557 (1989).
- Muller, J. *Bot. Rev.* **47**, 1–142 (1981).
- Palmer, J. D. *Trends Genet.* **6**, 115–120 (1990).
- Siemeister, G. & Hächter, W. *Curr. Genet.* **15**, 435–441 (1989).
- Siemeister, G., Buchholz, C. & Hächter, W. *Plant molec. Biol.* **14**, 825–833 (1990).
- Machado, M. A. & Zetsche, K. *Planta* **181**, 91–96 (1990).
- Kirk, J. T. O. & Tilney-Bassett, R. A. E. *The Plastids* (Elsevier/North-Holland, New York, 1978).
- Doyle, J. J. & Doyle, J. L. *Phytochem. Bull.* **19**, 11–15 (1987).
- Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning. A Laboratory Manual*. (Cold Spring Harbor Laboratory, New York, 1982).
- Palmer, J. D. *Meth. Enzym.* **118**, 167–186 (1986).
- Shinozaki, K. et al. *Plant molec. Biol. Rep.* **4**, 111–147 (1986).
- Hamby, R. K., Sims, L., Issel, L. & Zimmer, E. *Plant molec. Biol. Rep.* **6**, 175–192 (1988).
- Stern, D. B., Dyer, T. A. & Lonsdale, D. M. *Nucleic Acids Res.* **10**, 3333–3340 (1982).

ACKNOWLEDGEMENTS. We thank R. Eplee for fresh tissue of *Striga asiatica*, M. Sugiura for the parent tobacco clones from which the smaller ones used in this study were generated, and C. W. Morden and K. H. Wolfe for critical readings of an earlier draft of this paper. This study was supported by the NSF (C.W.D.) and the NIH (J.D.P.).

(Mg-ATP)-dependent self-assembly of molecular chaperone GroEL

N. M. Lissin, S. Yu. Venyaminov & A. S. Girshovich*

Institute of Protein Research of the Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

THE important *Escherichia coli* heat-shock protein GroEL of relative molecular mass 57,259 (ref. 1) is a typical molecular chaperone^{2,3}. It possesses ATPase activity^{4,5} and interacts in ATP-driven reactions with non-folded proteins^{6–8} to stimulate their correct folding and/or assembly by preventing the formation of improper protein structures or aggregates^{7,9–11}. As GroEL is isolated and functions as a 20–25S tetradecameric particle^{5,6,12} (GroEL_p), the question arises—what is the mechanism of its own assembly? Here we show the (Mg-ATP)-dependent self-stimulation ('self-chaperoning') *in vitro* of GroEL_p reassembly from its monomeric state.

We first studied the conditions of disassembly of GroEL_p. Gel filtration on Superose 6HR shows complete conversion of GroEL_p to its monomeric state (GroEL_m) after 20 min incubation on ice in the presence of 3.5 M urea. At 23 °C GroEL_p is stable against 3.5 M urea and its disassembly requires a higher (4 M) urea concentration. Gel filtration of isolated GroEL_m reveals no detectable GroEL_p, indicating that GroEL_p reassembly is not a spontaneous process. As GroEL_p is an ATPase, the effects of different adenine nucleotides on its reassembly at 23 °C were analysed. Table 1a shows that ATP plus Mg²⁺ (Mg-ATP), but not ATP or Mg²⁺ separately, are essential for effective reassembly of GroEL_p, the yield of which increases sigmoidally with increase in GroEL_m concentration up to about 75% at 7.7 µM GroEL_m (Fig. 1A). Reassembled GroEL_p is identical to native GroEL_p in its ATPase activity, sedimentation constant and microscopic appearance^{5,6,12} (Fig. 1B). Of note is the fact that in our disassembly-reassembly experiments GroEL_m and GroEL_p, but not intermediate species, are detected in gel-filtration profiles indicating that the oligomerization of GroEL_m is apparently highly cooperative. Reassembly must evidently proceed through some intermediate state(s) but the instability of the intermediate species precludes their detection by the method used.

(Mg-ATP)-dependent reassembly of GroEL_p shows a sharp dependence on temperature: incubation of GroEL_m on ice does not lead to detectable oligomerization (Table 1a). This implies that GroEL_m is converted on cooling into a state unable to form GroEL_p. The temperature dependence of the GroEL_m circular dichroism spectrum indicates that GroEL_m structure becomes disordered as temperature decreases ('cold denaturation'¹³).

* To whom correspondence should be addressed.

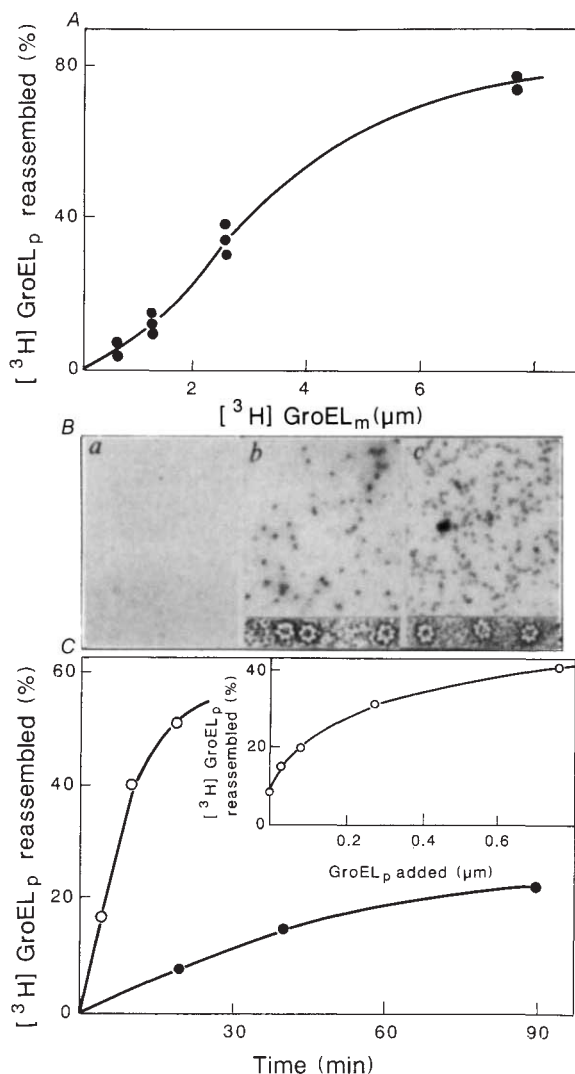
TABLE 1 Effects of additions on [^3H]GroEL_p* reassembly at 23 °C

	AMP	ADP	AMP-PNP	ATP- γ -S	ATP	Mg ²⁺	GroES	Reassembly of GroEL _p (%)
(a)	—	—	—	—	—	—	—	<0.3
	—	—	—	—	—	+	—	<0.3
	+	—	—	—	—	+	—	<0.3
	—	+	—	—	—	+	—	5.1
	—	—	+	—	—	+	—	8.0
	—	—	—	+	—	+	—	4.8
	—	—	—	—	+	+	—	35.6 (<0.3†)
	—	—	—	—	+	—	—	1.3
(b)	—	—	—	—	—	+	+	<0.3
	—	+	—	—	—	+	—	1.4
	—	+	—	—	—	+	+	1.6
	—	—	—	+	+	+	—	13.8
	—	—	—	+	+	+	+	55.0

2.6 μM (a) or 1.3 μM (b) [^3H]GroEL_m are incubated in buffer A (20 mM triethanolamine acetate, pH 7.5, 100 mM potassium-acetate, 1 mM dithiothreitol, 0.1 mM EDTA) with additions, where shown, of 1.7 mM adenine nucleotides, 5 mM magnesium acetate, or 1.9 μM GroES for 20 min at 23 °C. [^3H]GroEL_m and reassembled [^3H]GroEL_p are separated by chromatography on Superose 6HR 10/30 (Pharmacia FPLC-System) at 4 °C in buffer A (0.4 ml min⁻¹). GroEL_p and GroES are purified from *E. coli* HB101 bearing the multicopy plasmid pGroE4 which is constructed by cloning the *Eco*RI fragment of *E. coli* DNA, containing the complete *groES-groEL* operon, in the *Eco*RI site of vector plasmid pACYC 184. [^3H]GroEL_p is prepared by reductive methylation in the presence of [^3H]NaBH₄ (6.2 Ci mmol⁻¹, Amersham) and stored at -20 °C in buffer A containing 10 mM magnesium acetate; its specific activity is about 1,500 c.p.m. per pmol of GroEL_m. [^3H]GroEL_m is prepared by dilution of [^3H]GroEL_p (6 mg ml⁻¹) with 10 M urea to 4 M final concentration of the latter and incubation for 20 min at 23 °C with subsequent chromatography as described above. The isolated GroEL_m is stored on ice before use. The Superose column is calibrated at 23 °C and 4 °C using a Pharmacia gel-filtration calibration kit. The retention time (R_t) for GroEL_p (31 min) or for proteins of the kit does not change in this temperature region, whereas R_t of GroEL_m is 40 min at 23 °C (identical to R_t of BSA, M_r 67K) and decreases to 38.3 min at 4 °C (R_t 38 min for rabbit muscle aldolase, M_r 158K) apparently because of GroEL_m cold denaturation (Fig. 2). Non-hydrolysable ATP analogues: AMP-PNP, adenylyl-imidodiphosphate; ATP- γ -S, adenosine-5'-O-(3-thiotriphosphate).

* GroEL_p is used in a radioactive form to facilitate quantitative estimation of its yield after reassembly.

† Reassembly for 2 h in ice.



GroEL_m at room temperature is in a less ordered state than GroEL_p and is denatured not only with increasing temperature but, unlike GroEL_p, also on cooling (Fig. 2). Thus, the temperature decrease facilitates the urea-stimulated disassembly of GroEL_p and prevents its reassembly from the monomeric state. Of the various possible interactions, only hydrophobic interactions decrease with decreasing temperature¹⁴. It follows that hydrophobic interactions can have an important role in maintaining the oligomeric structure of GroEL_p.

The replacement of ATP by ADP or by ATP analogues that cannot be hydrolysed strongly decreases the yield of reassembled GroEL_p (Table 1). This indicates that ATP before its hydrolysis induces the slow GroEL_p reassembly perhaps by stabilization of a GroEL conformational intermediate that is essential for, or facilitates, the formation of the final oligomer. But a high efficiency of GroEL_p reassembly evidently requires ATP hydrolysis. The ATP dependence of GroEL_p reassembly correlates with the ATP requirement for stimulation of reassembly of the dimeric protein Rubisco (ribulose biphosphate carboxylase/oxidase)⁷ by GroEL_p and, in general, with the accepted principle of the ATP-driven functioning of molecular chaperones^{2,3,15-17}.

The following mechanism of 'self-chaperoning' of GroEL_p reassembly can be proposed: Mg-ATP, before its hydrolysis, initiates the reassembly of the primary GroEL_p which, in turn, acts as a chaperone, interacting with the remaining bulk of

FIG. 1 (Mg-ATP)-dependent reassembly of [^3H]GroEL_p at 23 °C. A, Effect of GroEL_m concentration on the yield of reassembled GroEL_p (20 min incubation); B, electron microscopy monitoring the course of GroEL_p reassembly (negative staining with uranyl acetate⁶). a, Zero-point; b and c, 2 and 10 min of GroEL_m incubation in the presence of Mg-ATP, respectively. Top, General fields; bottom, selected electron micrographs. C, Stimulation of the [^3H]GroEL_p reassembly in the presence of the native GroEL_p; concentrations of [^3H]GroEL_m and GroEL_p are 1.3 and 1.1 μM , respectively. ● or ○, The [^3H]GroEL_p reassembly without or in the presence of the native GroEL_p, respectively. Inset, Effect of added native GroEL_p concentration on the yield of the reassembled [^3H]GroEL_p (20 min incubation); concentration of [^3H]GroEL_m is 0.87 μM . Addition of Mg-ATP in buffer A, used as an eluent at isolation of the reassembled [^3H]GroEL_p by chromatography on Superose, does not affect the results.

GroEL_m and stimulating its oligomerization in an ATPase-dependent manner. To test this hypothesis, we imitated the initiation step of GroEL_p reassembly by addition of native non-radioactive GroEL_p. Indeed, its addition to the sample containing Mg-ATP and [³H]GroEL_m significantly increases the yield of [³H]GroEL_p (Fig. 1C). The latter is a true reassembled particle and not, for example, the [³H]GroEL_m·GroEL_p complex, as the radioactivity detected in GroEL_p does not decrease on gel filtration in the presence of ATP, which dissociates such complexes⁶. An *a priori* probability of (Mg-ATP)-dependent substitution of the GroEL_m component(s) of GroEL_p by [³H]GroEL_m, which could result in formation of a 'mixed' radioactive particle, must also be rejected. [³H]GroEL_m does not appear on incubation of [³H]GroEL_p with an excess of non-radioactive GroEL_m under conditions of incomplete conversion of the latter into GroEL_p (Fig. 3). Hence, the added GroEL_p indeed stimulates true GroEL_m oligomerization but not the exchange reaction. In accordance with the ability of GroEL_p to interact with one molecule of non-folded protein^{8,18}, the maximal rate of GroEL_m oligomerization is observed at ~1:1 molar ratio of GroEL_m to the added GroEL_p (inset, Fig. 1C).

The partner of GroEL_p required for its functioning in some cases is the *E. coli* heat-shock protein GroES (relative molecular mass 10,368 (ref. 1)). This exists as a heptameric particle¹⁹ which, in the presence of ATP, forms a 1:1 complex with GroEL_p (ref. 19) and helps it to stimulate assembly of some oligomeric proteins^{7,9,10}. Table 1b shows that GroES also stimulates the reassembly of GroEL_p. However, our case is a peculiar one as GroEL_p is simultaneously both a molecular chaperone and a reassembled protein. In particular, the (Mg-ATP)-dependent interaction with GroES could perhaps increase the yield of reassembled GroEL_p by stabilizing its oligomeric structure and thereby inhibiting its disassembly. But this version cannot be accepted as GroEL_p itself is stable on incubation with Mg-ATP as tested by gel filtration or by centrifugation in a sucrose

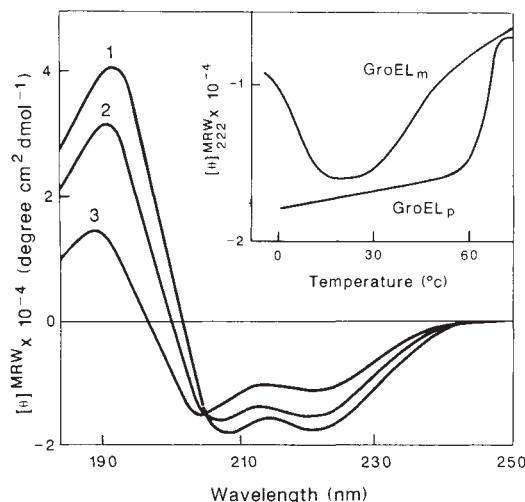


FIG. 2 Circular dichroism spectra of GroEL_p and GroEL_m. 1, GroEL_p at 20 °C or 0 °C; 2, GroEL_m at 20 °C; 3, GroEL_m at 0 °C. The relative amounts (%) of the different secondary structural elements determined according to ref. 21 from spectra 1, 2 and 3 are the following: α helix 57, 47 and 33; β sheet 30, 21 and 23; β turn 8, 8 and 7; remaining (non-regular) structure 5, 24 and 37, respectively. Inset, Temperature dependence of mean residue ellipticity at 222 nm of GroEL_p or GroEL_m. Transition points: GroEL_p, ~67 °C; GroEL_m, ~5 °C and ~43 °C.

METHODS. Circular dichroism measurements are made in buffer: potassium phosphate (50 mM), potassium acetate (20 mM), magnesium acetate (5 mM), dithiothreitol (0.1 mM), pH 7.5, using the Jasco-600 spectropolarimeter (Japan) equipped with a temperature-controlled cell holder. Protein concentration is measured spectrophotometrically at $\lambda_{\max}$ 276.5 nm; the absorbance values $A_{1\text{cm}}^{0.1\%}$ used of GroEL_p and GroEL_m are 0.25 and 0.21, respectively, determined by nitrogen analysis²².

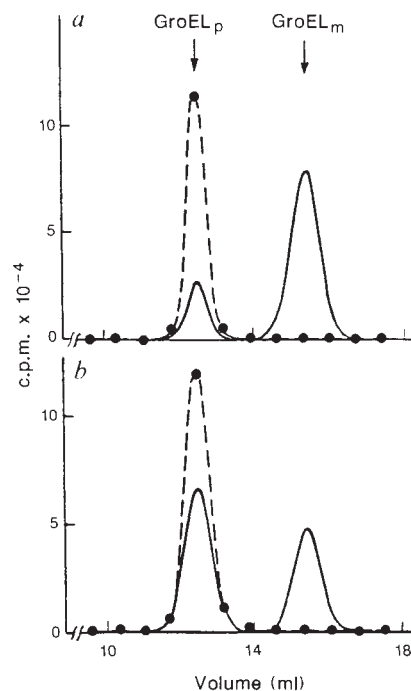


FIG. 3 The monomeric component of [³H]GroEL_p does not exchange with added GroEL_m.

METHODS. [³H]GroEL_p (0.05 μ M, about 0.7 μ M monomeric component) is mixed with 2.8 μ M GroEL_m in buffer A containing 1.7 mM ATP and 5 mM magnesium acetate and chromatographed on Superose, as described in the legend to Table 1, either immediately (a) or after 20 min incubation at 23 °C (b). —, 280 nm absorbance scanning; • — •, radioactivity. The incomplete GroEL_m oligomerization under the above conditions is confirmed by about a twofold increase of the GroEL_p reassembly yield on more prolonged incubation (1 h). Analogously, at [³H]GroEL_p and GroEL_m concentrations similar to those described in Fig. 1c, no radioactivity is detected at the GroEL_m position under conditions of about 40% conversion of GroEL_m into GroEL_p.

gradient or by non-denaturing PAGE, pH 7.4 (ref. 20) (not shown). Moreover, GroES is unable to stimulate the reassembly of GroEL_p in the presence of ADP (Table 1b) which stabilizes, similarly to ATP, the complex of GroEL_p with GroES (not shown) and prevents the chaperone function of GroEL_p. Thus, the revealed stimulatory effect of GroES (Table 1b) confirms the 'self-chaperoning' mechanism of GroEL_p reassembly. A logical consequence of this conclusion is that the need for GroES in order that GroEL_p should function in the cell now has an additional aspect—stimulation of assembly of this molecular chaperone by GroES.

It is evident that GroEL_p can effectively prevent incorrect folding or aggregation of cellular proteins *in vivo* only if the assembly of newly synthesized GroEL_m into the functionally active oligomer does not proceed later than these undesirable conformational conversions of the proteins. Therefore 'self-chaperoning', accelerating GroEL_p assembly, can be essential for the functioning of GroEL_p in the cell. In conclusion, the mechanism of the GroEL_p self-assembly shown here could be unique in nature as it is possible *a priori* only for an oligomeric protein which is a molecular chaperone. □

Received 20 July; accepted 26 September 1990.

- Hemmingsen, S. M. *et al.* *Nature* **333**, 330–334 (1988).
- Ellis, R. J. *Nature* **328**, 378–379 (1987).
- Ellis, R. J. & Hemmingsen, S. M. *Trends biochem. Sci.* **14**, 339–342 (1989).
- Ishihama, A., Ukeichi, T., Matsumoto, A. & Yamamoto, S. *J. Biochem., Tokyo* **79**, 927–936 (1976).
- Hendrix, R. W. *J. molec. Biol.* **129**, 375–392 (1979).
- Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254–257 (1988).
- Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **432**, 884–889 (1989).
- Lecker, S. *et al.* *EMBO J.* **8**, 2703–2709 (1989).
- Friedman, D. I., Olson, E. R., Georgopoulos, C., Tilly, K., Hershowitz, I. & Banuett, F. *Microbiol. Rev.* **48**, 299–325 (1984).
- Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. *Nature* **337**, 44–47 (1989).

11. Van Dyk, T. K., Gatenby, A. A. & LaRossa, R. A. *Nature* **342**, 451–453 (1989).
12. Hohn, T., Hohn, B., Engel, A., Wurtz, M. & Smith, P. R. *J. molec. Biol.* **129**, 359–373 (1979).
13. Privalov, P. L. *Rev. Biophys. biophys. Chem.* **18**, 47–69 (1989).
14. Privalov, P. L. & Gill, S. J. *Adv. Prot. Chem.* **39**, 191–324 (1988).
15. Pelham, H. R. B. *Cell* **46**, 959–961 (1986).
16. Rothman, J. E. *Cell* **59**, 591–601 (1989).
17. Kassenbrock, C. K. & Kelly, R. B. *EMBO J.* **8**, 1461–1467 (1989).
18. Laminet, A. A., Ziegelhoffer, T., Georgopoulos, C. & Pluckthun, A. *EMBO J.* **9**, 2315–2319 (1990).
19. Chandrasekhar, G. N., Tilly, K., Woolford, C., Hendrix, R. & Georgopoulos, C. *J. biol. Chem.* **261**, 12414–12419 (1986).
20. McLellan, T. *Analyt. Biochem.* **126**, 94–99 (1982).
21. Provencher, S. W. & Glockner, J. *Biochemistry* **20**, 33–37 (1981).
22. Jaenicke, L. *Analyt. Biochem.* **61**, 623–627 (1974).

ACKNOWLEDGEMENTS. We thank E. S. Bockkareva for helpful discussions, N. K. Osyna for GroEL electron microscopy and A. I. Gragorov for the generous gift of *E. coli* HB101(pGroE4).

Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region

Mary E. Reaban* & Johanna A. Griffin†

*Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

†Department of Molecular Biology, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, Connecticut 06877, USA

A DELETIONAL DNA rearrangement is associated with immunoglobulin class switching from IgM to IgG, IgA or IgE (reviewed in ref. 1). This recombination occurs in immunoglobulin switch regions, which are complex, highly repetitive regions of DNA. As switch regions become transcriptionally active^{2,3} just before switch recombination, analysis of the behaviour of these sequences during transcription could elucidate the mechanism of switch recombination. Here, we report that transcription of a supercoiled plasmid containing the murine IgA switch region ($S\alpha$) leads to a loss of superhelical turns. The resulting series of less supercoiled plasmids is stabilized by RNA–DNA hybrids formed by the nascent RNA transcripts, which remain base-paired with their DNA templates.

To study the interaction of transcription and immunoglobulin switch region DNA sequences, we constructed the plasmid pBS- α S2.3. A 2.3-kilobase (kb) *Pst*I–*Bam*HI restriction fragment from pS α 64101 (ref. 4) consisting of the murine IgA switch region was cloned into the vector pBS- at a site between the convergent T3 and T7 RNA polymerase promoters. Preliminary transcription studies revealed no polymerase stops or pauses that could be attributed to the presence of the switch region sequences, when supercoiled or linearized pBS- α S2.3 or pBS- plasmids were transcribed with either T3 or T7 polymerase (data not shown). Additionally, transcription did not significantly nick or break the supercoiled templates.

A DNA ladder of progressively less supercoiled plasmids was observed when supercoiled pBS- α S2.3 was electrophoresed after transcription with T7 RNA polymerase (Fig. 1, lane 3). The samples were treated with RNase A after transcription, to remove single-stranded RNA. All the supercoiled molecules were at least partially relaxed, with the amount of relaxation ranging from slight to complete. This phenomenon is a function of the transcribed sequence rather than of transcription, or of the individual transcribing polymerase. Transcription of the other strand of supercoiled pBS- α S2.3 with T3 RNA polymerase (Fig. 1, lane 2) did not have this effect, nor did T3 nor T7 transcription of pBS- (Fig. 1, lanes 8 and 9). Furthermore, transcription of linearized plasmids with either polymerase did not result in any aberrantly migrating species (Fig. 1, lanes 5, 6, 11, 12). As formation of the ladder of plasmid bands was only seen following T7 transcription of supercoiled pBS- α S2.3, the transcription of a specific strand of the switch region insert resulted in relaxed supercoils, perhaps because of the formation of a supercoil-relaxing structure.

A persistent RNA is responsible for the appearance of the relaxed plasmids. When T7-transcribed pBS- α S2.3 was treated with RNase H, the plasmids that had previously formed the ladder of bands (Fig. 2, lane 6) migrated with the native supercoiled form of the plasmid (Fig. 2, lane 7). This result indicated that an RNA–DNA hybrid had stabilized the less supercoiled plasmids, because RNase H specifically degrades the RNA in an RNA–DNA hybrid. Additionally, this reversion of plasmids to their original supercoiled state showed that the formation of the ladder was due to a conformational change in the plasmids rather than to a change in their linking number. Thus, the less supercoiled plasmids are conformers rather than topoisomers. Also, the transcription-generated conformers clearly all contained a single-stranded region of DNA, because treatment with mung bean nuclease (specific for single-stranded DNA and RNA) caused them to migrate with the nicked form of the plasmid (Fig. 2, lane 8). If the conformers were treated with RNase H, they lost their sensitivity to mung bean nuclease (Fig. 2, lane 9), which indicated that an RNA–DNA hybrid generated or enhanced the single-stranded DNA region.

The RNA transcript alone was insufficient to induce the formation of the conformers under transcription conditions. T3 or T7 transcripts made from uncut pBS- α S2.3 or pBS- plasmids were purified from their templates, then added back to uncut plasmids. The mixtures were incubated under conditions identical to those used for transcription, except that no RNA polymerase was included. When these samples were electrophoresed, no ladder of relaxed plasmids was seen (data not shown). This implies that, in addition to the RNA transcript, transcriptional activity was also required to form the conformers.

The translocation of an RNA polymerase during transcription may generate positive supercoils in the DNA ahead of the polymerase and negative supercoils behind it⁵. Studies conducted *in vitro*⁶ and *in vivo*^{7,8} support this twin supercoil-domain model. The localized negative supercoiling thus generated in the wake of the polymerase has been postulated to be sufficient

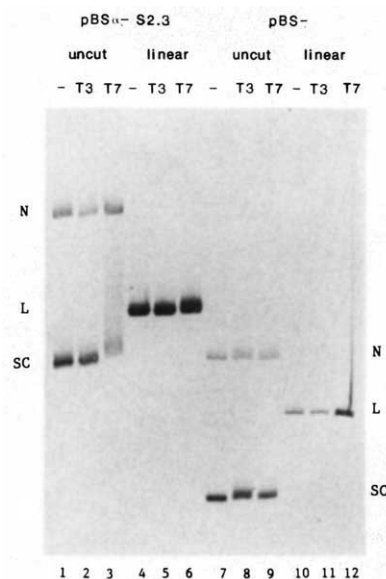


FIG. 1 Effect of transcription of the IgA switch region. Uncut and linear plasmids (5×10^{10} copies) were transcribed for 20 min using either T3, T7, or no (–) polymerase. Transcription conditions were: 40 mM Tris buffer, pH8, 8 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 30 mM DTT, 0.6 mM each ATP, CTP, GTP and UTP, and 6 units polymerase (International Biotechnologies); total volume 25 μ l. The reactions were terminated by extraction with 1:1 phenol:CHCl₃ and treated with boiled DNase-free RNase A (1 μ g, Calbiochem) for 1 h at 37 °C to remove single-stranded RNA. Electrophoresis was performed in a TBE-buffered 1.2% agarose gel, and the DNA was visualized under ultraviolet light with ethidium bromide. N, nicked plasmid; L, linear plasmid; SC, supercoiled plasmid. Plasmids relaxed with wheatgerm topoisomerase I (Promega) co-migrated with the nicked molecules (data not shown).

TABLE 1 Densitometric analysis of conformer distribution

Band*	Area	Percent total area	Supercoils†	Supercoils relaxed
a	0.639	36.2	0	19
b	0.013	0.7	1	18
c	0.015	0.8	2	17
d	0.016	0.9	3	16
e	0.025	1.4	4	15
f	0.052	2.9	5	14
g	0.074	4.2	6	13
h	0.067	3.8	7	12
i	0.047	2.6	8	11
j	0.013	0.7	9	10
k	0.017	1.0	10	9
l	0.014	0.8	11	8
m	0.011	0.6	12	7
n	0.012	0.7	13	6
o	0.013	0.7	14	5
p	0.019	1.1	15	4
q	0.726	41.0	16-19	3-0

* The ladder of bands in lane 6 of Fig. 2 was scanned with an LKB Ultrascan XL laser densitometer and analysed using LKB gelscan XL software. The bands are labelled alphabetically in order of increasing number of supercoils (from top to bottom in the photograph).

† Nontranscribed pBS- α S2.3 was estimated to contain ~19 unrestrained supercoils by two-dimensional electrophoretic analysis as described in ref. 9.

to drive DNA structural transitions. We had previously found that the polypurine-polypyrimidine repeat (AGGAG)₂₈ contained in the α insert in pBS- α S2.3 forms an intramolecular DNA triplex under negative supercoiling conditions⁹. As the transition from normal B-DNA to such alternative structures relaxes supercoils, the stable relaxed conformers discussed here could be related to the formation of the intramolecular triplex. We suggest that there are two consequences of transcription of α . First, it transiently increases the level of negative supercoiling behind the RNA polymerase, which causes the (AGGAG)₂₈ repeat to assume its intramolecular triplex form. Second, it generates the complementary RNA transcript, which then stabilizes the triplex by base-pairing with the region of single-stranded DNA generated by its formation (Fig. 3). The net result would be a series of partially relaxed plasmids, such as we have seen.

The strand-specific transcriptional relaxation data (Fig. 1) support this idea, as only the polypyrimidine RNA transcript produced by T7 polymerase would be complementary to the single-stranded polypurine region generated by the formation

FIG. 3 Proposed post-transcription α structure. The polypurine-polypyrimidine repeat (AGGAG)₂₈ in pBS- α S2.3 forms an intramolecular DNA triplex at a superhelical density of -0.02 (ref. 9). In the triplex, the 3' half of the purine strand is simultaneously joined to the 3' half of the pyrimidine strand through Hoogsteen basepairs ($\wedge$) and to the 5' half of the pyrimidine strand through Watson-Crick basepairs (-). This configuration leaves the 5' half of the polypurine strand uninvolved in the triplex and unpaired. The polypyrimidine-containing RNA transcribed by T7 RNA polymerase has the ability to base-pair with the single-stranded polypurine region, and could thus stabilize the supercoil-relaxing intramolecular DNA triplex. Although there are 28 AGGAG repeats in α , only six are depicted here. X, bases not participating in triplex formation.

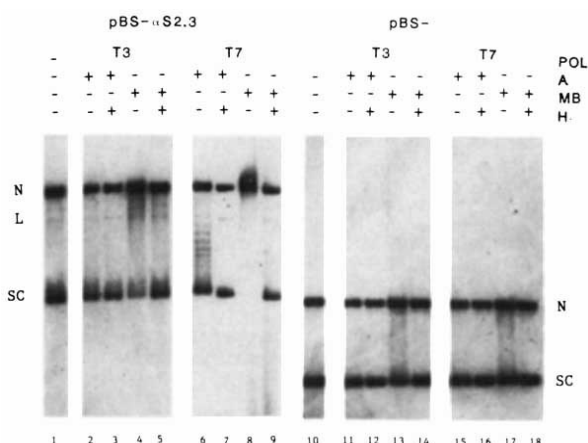
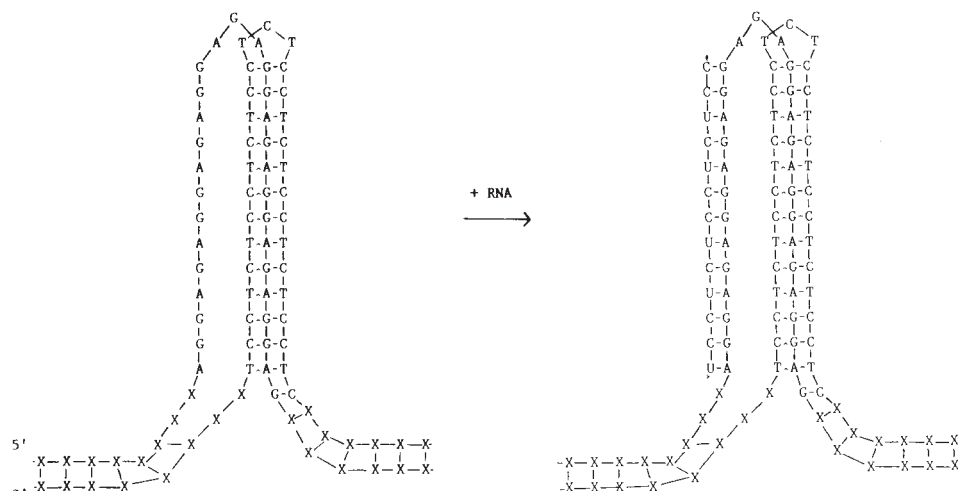


FIG. 2 Nuclease treatment of transcription products. Uncut plasmids were transcribed and extracted as described (Fig. 1). Aliquots of each sample were treated with either RNase A (A, 3 μ g) or mung bean nuclease (MB; 10 U, New England Biolabs) in the presence or absence of RNase H (H; 1.24 U, Boehringer-Mannheim) for 3 h at 37 °C. The pH at which this experiment was performed was not optimal for mung bean nuclease, therefore RNase H was much more efficient. Electrophoresis was as described (Fig. 1). POL, transcribing polymerase.

of the triplex. It is interesting to note that this direction of transcription is the same as that used *in vivo* to produce the pre-switch α transcripts¹⁰.

The amount of transcriptionally induced supercoil relaxation is also consistent with the formation and stabilization of the intramolecular triplex. The most intense band in the conformer ladder correlates with plasmids that have had 13 supercoils relaxed, followed closely by the species with 12 supercoils relaxed (Table 1). The formation of an intramolecular triplex by the (AGGAG)₂₈ repeat is predicted to relax 12.73 supercoils¹¹. Some of the conformers exhibit a greater degree of relaxation than can be attributed to the triplex-forming region alone. It is conceivable that the triplex serves as a site of nucleation for RNA-DNA hybridization, which then extends to relax the plasmid completely.

We have shown that transcription of the polypurine-containing strand of the IgA switch region in a supercoiled plasmid results in the formation of stable, less supercoiled conformers of the plasmid. The conformers are stabilized by the nascent RNA transcripts, which remain associated with their templates through RNase H-sensitive RNA-DNA hybrids. The formation

of such RNA-DNA hybrids in transcriptionally active switch regions may be important in switch recombination. □

Received 24 August; accepted 10 October 1990.

1. Lutzker, S. G. & Alt, F. W. in *Mobile DNA* (eds Howe, M. M. & Berg, D. E.) (American Society for Microbiology, Washington DC, 1989).
2. Stavnezer, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7704-7708 (1988).
3. Lutzker, S., Rothman, P., Pollock, R., Coffman, R. & Alt, F. W. *Cell* **53**, 177-184 (1988).
4. Stanton, L. W. & Marcu, K. B. *Nucleic Acids Res.* **10**, 5993-6006 (1982).
5. Liu, L. F. & Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7024-7027 (1987).
6. Tsao, Y.-P., Wu, H.-Y. & Liu, L. F. *Cell* **56**, 111-118 (1989).
7. Wu, H.-Y., Shyy, S., Wang, J. C. & Liu, L. F. *Cell* **53**, 433-440 (1988).
8. Giaever, G. N. & Wang, J. C. *Cell* **55**, 849-856 (1988).
9. Collier, D. A., Griffin, J. A. & Wells, R. D. *J. biol. Chem.* **263**, 7397-7405 (1988).
10. Radcliffe, G., Lin, Y.-C., Julius, M., Marcu, K. B. & Stavnezer, J. *Molec. cell. Biol.* **10**, 382-386 (1990).
11. Htun, H. & Dahlberg, J. E. *Science* **243**, 1571-1576 (1989).

ACKNOWLEDGEMENTS. We thank K. Marcu for pS α 64101, and P. Higgins, G. Knapp and J. Lebowitz for comments. This work was supported by the NIH (J.A.G.), and the National Cancer Institute (M.E.R.). J.A.G. was an established investigator of the American Heart Association.

Mediation of virion penetration into vascular cells by association of basic fibroblast growth factor with herpes simplex virus type 1

Andrew Baird, Robert Z. Florkiewicz,
Pamela A. Maher, Robert J. Kaner*
& David P. Hajjar†

Department of Molecular and Cellular Growth Biology,
The Whittier Institute For Diabetes and Endocrinology, La Jolla,
California 92037, USA

* Pulmonary Service, Department of Medicine, Memorial-Sloan Kettering
Cancer Center, New York, New York 10021, USA

† Departments of Biochemistry and Pathology,
Cornell University Medical College, New York, New York 10021, USA

HERPES simplex virus type-1 (HSV-1) is a ubiquitous pathogen that is associated with considerable morbidity in the general population. Although it is known that the virion uses a basic fibroblast growth factor (FGF) receptor to penetrate vascular cells¹, it is not known how the viral particle recognizes and binds to this cell surface protein. Here we report that an immunoreactive basic FGF-like protein is associated with the viral particle and that this association appears responsible for viral uptake. Accordingly, HSV-1 infection of Swiss 3T3 cells stimulates the tyrosine phosphorylation of the specific substrate that characterizes the initial cellular response to basic FGF. Antibodies to basic FGF prevent this phosphorylation and inhibit HSV-1 uptake. Because no basic FGF sequence is found in the HSV-1 genome, a model for the infection for some target cells is presented whereby the viral particle uses host cell-derived basic FGF to ensure subsequent infectivity of newly replicated virus.

Recombinant basic FGF inhibits the uptake and infectivity of HSV-1 because the virion uses the basic FGF receptor to penetrate vascular cells¹. To examine the possibility that HSV-1 and basic FGF compete for the same sites on high and low affinity receptors, we performed radioreceptor assays². An HSV-1 preparation displaces the binding of radiolabelled basic FGF to its high affinity receptor (Fig. 1) and prevents the detection of the crosslinked ¹²⁵I-labelled basic FGF-receptor complex (inset to Fig. 1).

HSV-1 also activates the high affinity basic FGF receptor (Fig. 2). Incubation of Swiss 3T3 fibroblasts with HSV-1 results in the tyrosine phosphorylation of a unique protein substrate of relative molecular mass 90,000 (*M*_r 90 K) that characterizes the cellular response to basic FGF (refs 3, 4). The DNA virus,

adenovirus, has no effect. Because an immunoneutralizing antibody to basic FGF (ref. 5) inhibits the HSV-1-dependent phosphorylation of the 90 K protein (Fig. 2), it seems that HSV-1 either releases basic FGF from target cells or that a basic FGF-like protein is associated with the viral particle.

There are no known proteins encoded in the HSV-1 genome that have significant (>25%) structural homology with fragments of basic FGF (ref. 1). Thus, it is difficult to account for the virion's capacity to bind and activate the basic FGF receptor. However, immunoneutralizing antibodies to basic FGF (ref. 5) can prevent viral uptake into target cells (Fig. 3), suggesting that a basic FGF-like epitope exists on the virion surface. Several viruses have been shown to use the binding sites for known physiological ligands as portals of entry into target cells and it is not always known how they recognize these receptors. As an example, rhinovirus uses the receptor for the cell adhesion molecule ICAM-1 to penetrate cells^{6,7} and the HIV virus uses the CD4 glycoprotein receptor^{8,9}. The Epstein-Barr virus infects T lymphocytes by the C3d complement receptor¹⁰, the rabies virus infects cells through the acetylcholine receptor¹¹, the rheovirus enters through the β -adrenergic receptor¹², and the vaccinia virus enters the cell by first interacting with the epidermal growth factor (EGF) receptor¹³. In the latter instance, the vaccinia virus genome encodes a protein (VVGf) that is highly homologous to EGF and that, in turn, allows the viral particle to recognize the EGF receptor¹³. A similar, yet distinct, mechanism seems to allow HSV-1 to recognize the basic FGF receptor.

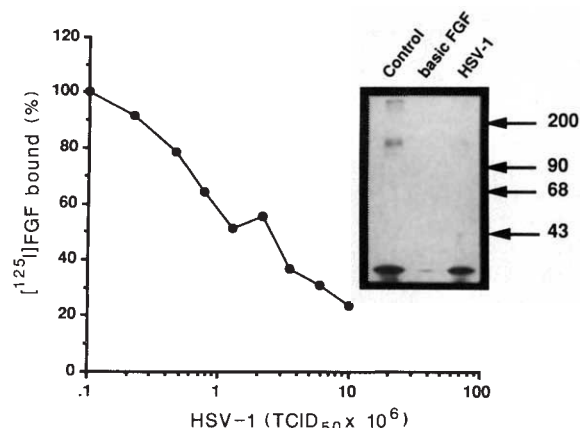


FIG. 1 Effect of HSV-1 on the binding of radiolabelled basic FGF to its receptor on baby hamster kidney cells. Various concentrations of purified HSV-1 particles (type 1, strain F from facial vesicle, expressed as tissue culture infective dose, TCID₅₀) were incubated with 600 pg of ¹²⁵I-labelled basic FGF (60,000 c.p.m.) on subconfluent baby hamster kidney cells for 3 h at 4 °C in 300 μ l M199 media. The cells were then washed once with PBS, twice with PBS containing 2 M NaCl to remove radiolabelled basic FGF bound to cell-associated glycosaminoglycans² and then the cells were solubilized with 0.5 ml of 0.1% Triton X-100 in PBS (pH 8.5). The solutions were counted and normalized to the maximal amounts of basic FGF binding. Nonspecific binding (7%) was determined by incubating the cells with 4 μ g ml⁻¹ of basic FGF and the values were subtracted from the counts before normalization. In contrast to the results obtained using unlabelled basic FGF, there was no detectable displacement of ¹²⁵I-labelled basic FGF by HSV-1 detected in any of the 2 M washes (that is, GAG binding) indicating that the two ligands are not competing for the low affinity receptor. Whereas 95% of the labelled basic FGF was displaced from the low affinity sites by basic FGF (100 ng ml⁻¹), <10% was displaced by 10⁷ TCID₅₀ HSV-1. Inset: ¹²⁵I-labelled basic FGF (10 ng ml⁻¹) was incubated for 3 h with subconfluent BHK cells grown in a 10-cm tissue culture dish. The incubation media contained either no additives (Control), basic FGF (4 μ g ml⁻¹) or HSV-1 (10⁶ TCID₅₀ ml⁻¹). The ¹²⁵I-labelled basic FGF was then cross-linked to its receptor using disuccinimidyl suberate^{25,26} the cells washed with PBS and 2 M NaCl in PBS as described above, and the detergent (0.1% Triton X-100) cell lysates were boiled in Laemmli's loading buffer²⁷. Identical amounts of protein were analysed by SDS-PAGE under reducing conditions and the radiolabelled receptor (~150K) was visualized by autoradiography. *M*_r markers are indicated on the right (K). The experiment was repeated twice with near-identical results.

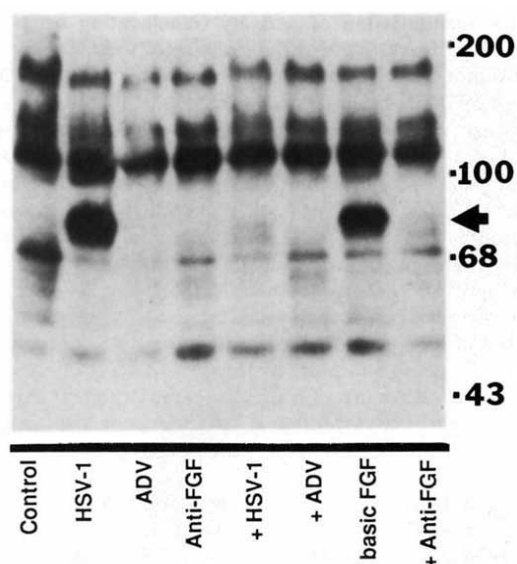


FIG. 2 Stimulation of protein tyrosine phosphorylation by HSV-1. Swiss 3T3 cells were treated at 37 °C for 5 minutes with no additives or with basic FGF (15 ng ml⁻¹), HSV-1 (5 × 10⁶ TCID₅₀) or adenovirus (5 × 10⁶ plaque forming units). In some instances (labelled +), the same additives were also tested after a 15 min preincubation with a 1:100 final dilution of anti-basic FGF antibody⁵. The cells were then collected in a 2.5 × Laemmli's buffer²⁷, the proteins separated on 8% SDS-PAGE gels and the presence of tyrosine phosphorylated proteins examined by western blotting with a specific anti-phosphotyrosine antibody^{28,29}. The *M_r* markers (K) are indicated on the right and the arrow highlights the 90K protein that is characteristically tyrosine phosphorylated in response to basic FGF. The experiment was repeated twice with near-identical results.

Western blot analyses of density-gradient-purified virus¹⁴ established the presence of an 18K immunoreactive protein that is indistinguishable from human recombinant basic FGF (Fig. 4). Because lysates of mock-infected cells or of purified adenovirus are devoid of any similar protein (not shown), it is possible to explain the ability of HSV-1 to recognize the basic FGF receptor by having basic FGF (of presumably cellular origin) associated with the virion.

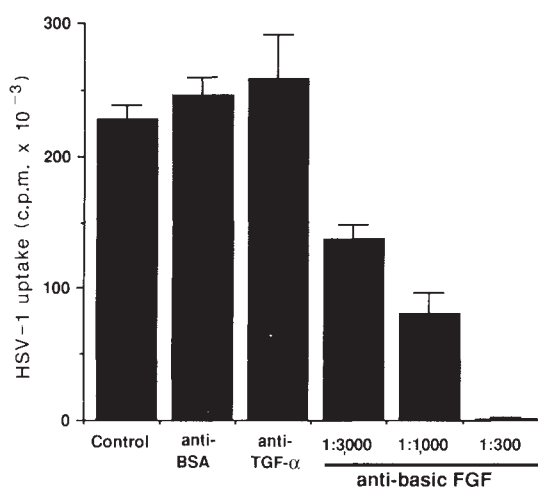


FIG. 3 Inhibition of HSV-1 uptake by antibodies to basic FGF. CHO cells, transfected with the basic FGF receptor *flg* gene³⁰ were inoculated with [³H-Tdr]-labelled HSV-1 at 37 °C for 2 h. Labelled virus was prepared as described^{1,31}. An aliquot of the virus was preincubated with control antibodies (anti-BSA, anti-TGF-α) or with the indicated concentrations of anti-basic FGF^{5,32}. To determine the uptake of virus, the cells were washed twice with PBS and the amount of radioactivity that was cell-associated was determined by scintillation counting^{1,31}. Bars represent the mean of quadruplicate determinations.

The mechanism through which a host cell-derived basic FGF becomes associated with the replicating viral particle is not known. It is most likely that the virus associates with the growth factor at the time of virion budding from the nuclear membrane as this is one of the locations where the 18K form of basic FGF has been localized¹⁵. We propose that before viral shedding, there is an association between the host cell-derived basic FGF and the newly synthesized HSV-1 virion that ultimately gives the viral particle the capacity to interact with the basic FGF receptor on other target cells. Thus, HSV-1 is like vesicular stomatitis virus and can incorporate into the virion host cell surface proteins¹⁶. Perhaps, more importantly, the results presented here suggest that HSV-1, like cytomegalovirus¹⁷, uses these target cell-derived proteins to confer infectivity. Like CMV however^{18,19}, the association of the receptor's ligand with the virion particle may not be the sole mechanism of receptor binding and internalization. For this reason, it will be particularly important to establish whether the addition of basic FGF to inactivated and/or attenuated virus can restore infectivity. Similar studies with HSV-2 are underway.

As herpesvirus latency is promoted by nerve growth factor *in vitro*,²⁰ it is interesting that the same virus uses basic FGF, another brain-derived neurotrophic factor²¹, for cell penetration. How these two characteristics of HSV-1 influence its potent neurotropism remains to be established. Because HSV-1 is ubiquitous, causing morbidity for the general population and mortality for the immunocompromised host, the identification of the mechanisms responsible for its entry into cells is of importance in developing strategies to prevent or alter the course of diseases induced by herpesvirus. The identification of basic FGF associated with the viral particles further validates the use of basic FGF antagonists¹ to prevent viral uptake and infectivity. Although the availability of these specific competitive²² and noncompetitive²³ receptor antagonists may help in the development of novel therapeutic strategies to control HSV-1 infection¹,

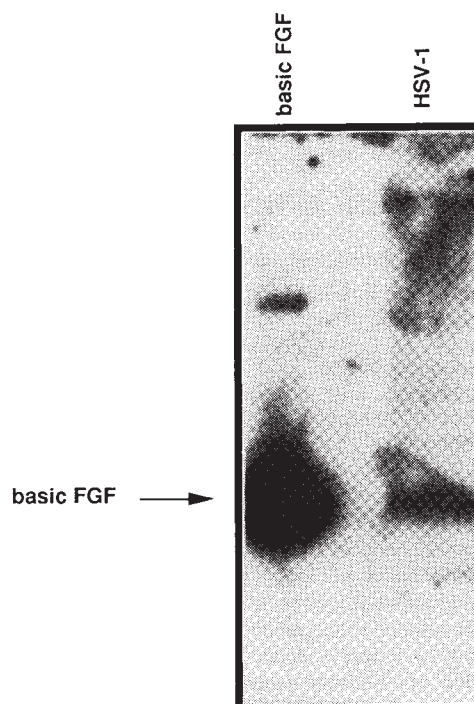


FIG. 4 Detection of basic FGF in purified HSV-1 preparations. A total of 10⁸ plaque forming units from infected cells were purified by density gradient centrifugation, concentrated, solubilized in Laemmli's sample buffer²⁷ and analysed on 15% SDS-PAGE reducing gels. The proteins were transferred to nitrocellulose and a western blot was performed using anti-basic FGF^{5,32}. The migration of recombinant human basic FGF (20 ng) is shown.

it is still important to establish the extent to which viruses use cell-derived ligands to increase their uptake by and infection of target cells and establish the importance of these pathways *in vivo*. The ability of a viral particle to use more than one cell-derived ligand might represent a mechanism for the infectious particle to avoid detection²⁴, to increase its virulence¹⁷, and potentially alter its cell and tissue tropism. □

Received 12 July; accepted 13 September 1990.

1. Kaner, R. J. *et al. Science* **248**, 1410–1413 (1990).
2. Moscatelli, D. J. *cell. Physiol.* **131**, 123–130 (1987).
3. Coughlin, S. R., Barr, P. J., Cousens, L. S., Fretto, L. J. & Williams, L. T. *J. biol. Chem.* **263**, 988–993 (1988).
4. Pasquale, E. B., Maher, P. A. & Singer, S. J. *J. cell. Physiol.* **137**, 146–156 (1988).
5. Gonzalez, A. M., Buscaglia, M., Ong, M. & Baird, A. *J. cell Biol.* **110**, 347–358 (1990).
6. Greve, J. M. *et al. Cell* **56**, 839–847 (1989).
7. Staunton, D. E. *et al. Cell* **56**, 849–853 (1989).
8. Dalglish, A. G. *et al. Nature* **312**, 763–764 (1984).
9. Klatzmann, D. *et al. Nature* **312**, 767–769 (1984).
10. Fingerth, J. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 4510–4514 (1984).
11. Lentz, T. L., Burrage, T. G., Smith, A. L., Crick, J. & Tignor, G. H. *Science* **215**, 182–184 (1982).
12. Man Sung Co, Gaulton, G. N., Fields, B. N. & Greene, M. I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1494–1498 (1985).
13. Eppstein, D. A. *et al. Nature* **318**, 663–665 (1985).
14. Cassai, E. N., Sarmiento, M. & Spear, P. G. *J. Virol.* **16**, 1327–1335 (1975).
15. Renko, M., Quarto, N., Morimoto, D. B., Rifkin, D. B. *J. cell. Physiol.* **144**, 108–114 (1990).
16. Lodish, H. F. & Porter, M. *Cell* **19**, 161–169 (1980).
17. Grundy, J. E., McKeating, J. A., Ward, P. J., Sanderson, A. R. & Griffiths, P. P. *J. gen. Virol.* **68**, 793–803 (1987).
18. Keay, S., Merigan, T. C., Rasmussen, L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10100–10103 (1989).
19. Taylor, H. P. & Cooper, N. R. *J. Virol.* **64**, 2484–2490 (1990).
20. Wilcox, C. L. & Johnson, E. M. Jr *J. Virol.* **61**, 2311–2315 (1987).
21. Wallike, P. A. & Baird, A. *Prog. Brain Res.* **78**, 333–338 (1988).
22. Baird, A., Schubert, D., Ling, N. & Guillemin, R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2324–2328 (1988).
23. Lappi, D. A., Martineau, D. & Baird, A. *Biochem. Biophys. Res. Commun.* **160**, 917–923 (1989).
24. McKeating, J. A., Griffiths, P. P. & Grundy, J. E. *J. gen. Virol.* **68**, 785–792 (1987).
25. Neufeld, G. & Gospodarowicz, D. *J. biol. Chem.* **260**, 13860–13868 (1985).
26. Olwin, B. B. & Hauschka, S. D. *Biochemistry* **25**, 3487–3492 (1986).
27. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
28. Maher, P. A. & Pasquale, E. B. *J. Cell Biol.* **106**, 1747–1755 (1988).
29. Wang, J. Y. *J. Molec. cell. Biol.* **5**, 3640–3643 (1985).
30. Mansukhani, A., Moscatelli, D., Talarico, D., Levyska, V. & Basilico, C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4378–4382 (1990).
31. WuDunn, D. & Spear, P. J. *J. Virol.* **63**, 52–58 (1989).
32. Halaban, R. *et al. Ann. N.Y. Acad. Sci.* **548**, 180–190 (1988).

ACKNOWLEDGEMENTS. Supported by the National Institutes of Health (A.B., D.P.H.), (P.A.M.), New York Heart Association Grant in Aid (D.P.H.) and the Erbman-Whittier Institute Angiogenesis Research Program (R.Z.F., P.A.M.). We thank J. Sharkey and M. Ong for technical assistance.

Force generation of organelle transport measured *in vivo* by an infrared laser trap

A. Ashkin*, Karin Schütze†, J. M. Dziedzic*, Ursula Euteneuer† & Manfred Schliwa†

* Laser Science Research Department, AT&T Bell Laboratories, Holmdel, New Jersey 07733, USA

† Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

ORGANELLE transport along microtubules is believed to be mediated by organelle-associated force-generating molecules¹. Two classes of microtubule-based organelle motors have been identified: kinesin^{2–7} and cytoplasmic dynein^{8–12}. To correlate the mechanochemical basis of force generation with the *in vivo* behaviour of organelles, it is important to quantify the force needed to propel an organelle along microtubules and to determine the force generated by a single motor molecule. Measurements of force generation are possible under selected conditions *in vitro* (for example, see refs 13 and 14), but are much more difficult using intact or reactivated cells. Here we combine a useful model system for the study of organelle transport, the giant amoeba *Reticulomyxa*¹⁵, with a novel technique for the non-invasive manipulation of and force application to subcellular components, which is based on a gradient-force optical trap, also referred to as 'optical tweezers'^{16–19}. We demonstrate the feasibility of using

controlled manipulation of actively translocating organelles to measure direct force. We have determined the force driving a single organelle along microtubules, allowing us to estimate the force generated by a single motor to be 2.6×10^{-7} dynes.

We measured the driving force of mitochondria moving along microtubules within fine strands of the peripheral network of *Reticulomyxa*. These mitochondria are spherical, with a uniform diameter of 320 ± 70 nm ($n = 50$) by electron microscopy; they have a high refractive index and move in large numbers at relatively high speed within strands, being clearly distinguishable by light microscopy from other particles. The number of motor molecules per mitochondrion is small, between 1 and 4. This is based on electron microscopy of microtubule-mitochondria complexes showing an average of 2.4 ± 1 ($n = 77$) cross-bridges and a maximum of four crossbridges per organelle (Fig. 1). The motors in *Reticulomyxa* are thought to be cytoplasmic dynein^{11,20} and to be capable of transport in opposite directions along microtubules^{11,20–22}.

To trap a mitochondrion moving along a strand requires a maximum trapping force greater than the maximum driving force, irrespective of the viscosity of the medium. A conceptually simple way of measuring the force is to trap a particle at high power and gradually reduce it until the particle just escapes the trap. This 'escape' power is a measure of the force. In practice this is difficult because of the heavy traffic of particles in active strands, so we used another approach based on limits. The procedure was to catch any actively moving single mitochondrion at an initially high power, and then to reduce the power quickly to a preset lower value to see whether it could escape. The force of any escaping particle has a value (equating, for the moment, power and force) between the upper and lower limits fixed by the stopping and escape powers.

Even at high power, some particles can elude the trap owing to the small size of the trap²³ relative to the size of the strand. Trapping is best in narrow, freely suspended strands $0.3 \mu\text{m}$ or less in diameter, when the trap captures the strand itself, which then guides mitochondria to the trap. In such narrow strands, mitochondria ride on the outside of a bundle of 1–6 microtubules, bulging the plasma membrane as they travel²⁴.

Figure 2 shows some simple experiments involving trapping and release of mitochondria moving along strands. These experiments illustrate the ability to trap fast-moving organelles and show that there are no obvious differences in behaviour of organelles moving in opposite directions along strands. Trap and release experiments can also roughly locate the relevant power (force) range corresponding to molecular motors. As we lower the power, the number of mitochondria trapped begins to drop at ~ 220 mW and is essentially zero at 30 mW, indicating that most mitochondria have forces between these power extremes. Table 1 summarizes the 'trap and escape' force measurements on 30 particles, for which the upper limit is taken at 220 mW and the lower limit varies from 30 to 110 mW, well into the relevant range. We list the number of particles measured in each of two power groups and give the average speeds measured before trapping and after escape. The fact that there

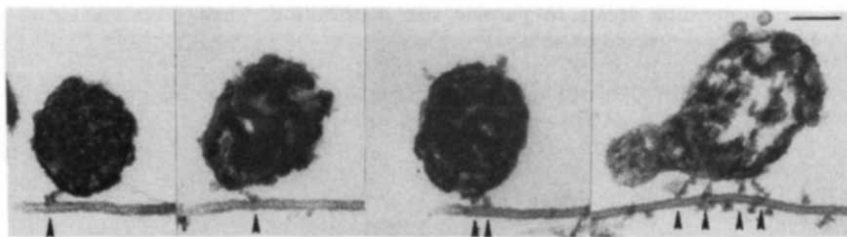
TABLE 1 Particle behaviour in 'trap and escape' experiments

Power range at escape (mW)	N	Average speed before trapping ($\mu\text{m s}^{-1}$)	Average speed after trapping ($\mu\text{m s}^{-1}$)
30–80	13	14.9	11.6
80–110	17	10.7	8.8

N is the total number of mitochondria observed in escape experiments and analysed for speed before trapping and after escape. All particles were trapped at 220 mW. The power was then dropped rapidly to a preset lower value by rotating the attenuator against a mechanical stop. Observations are from at least 12 strands in 3 amoeba samples.

FIG. 1 Examples of mitochondria with 1, 2 and 4 crossbridges to a microtubule, as seen by electron microscopy. Bar, 0.1 μm . To avoid confusion of *bona fide* crossbridges with other components of the ground cytoplasm, we used cells lysed with Brij 58 in a stabilization buffer³². After this gentle lysis, movement of mitochondria along microtubules can be reactivated by ATP (ref. 20), indicating that the mitochondria have active crossbridges to microtubules.

METHODS. Networks spread on coverslips were lysed in a buffer consisting of 50% PHEM (ref. 32; 60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM MgCl_2) 0.2% Brij 58, 1 mM vanadate and 5% hexylene glycol²⁰. After 1–2 min they were fixed with 0.5% glutaraldehyde



in 50% PHEM buffer and processed for electron microscopy³³. Sections were cut on a Porter-Blum MT2b microtome and viewed in a JEOL 1200 CX electron microscope.

is no significant change in speed after trapping indicates that the motors are not being damaged by the laser light. For a given particle that was, for example, initially trapped at 220 mW and later escaped at 100 mW, the force is taken as $(220 + 100)/2 = 160$ mW, the value midway between the upper and lower limits. Averaging over the 30 particles of Table 1 gives a mean driving force of ~ 150 mW per mitochondrion. Assuming 2.4 motors per mitochondrion (the value determined from electron microscopy), an average driving force for a single motor of ~ 63 mW is obtained.

The force of a trap on a particle depends on its size and relative index of refraction²³. To calibrate the force inside the amoeba, we compared the viscous drag forces on mitochondria in water with those in the amoeba. Free mitochondria were

obtained by gentle cell homogenization in 100 mM HEPES buffer. From the drag velocity at which mitochondria just escape, we deduce (using Stoke's law) that 1.0 mW gives a maximum trapping force of 5.8×10^{-9} dynes ($n = 21$) in water. In a comparable experiment within large low-viscosity amoeba droplets, it takes about twice the power for the same maximum drag velocity. We attribute this factor of two to either decreased force (due to a presumed lower relative index of refraction) or increased viscosity in the amoeba. By assuming a force $\sqrt{2}$ smaller and viscosity $\sqrt{2}$ higher, we make a maximum error of $\sqrt{2}$ in these quantities. On this basis, 1.0 mW gives a force of 4.1×10^{-9} dynes inside the amoeba. The value of 63 mW for the force of one motor translates into 2.6×10^{-7} dynes for the force generated by a single motor molecule, with an overall accuracy of a factor

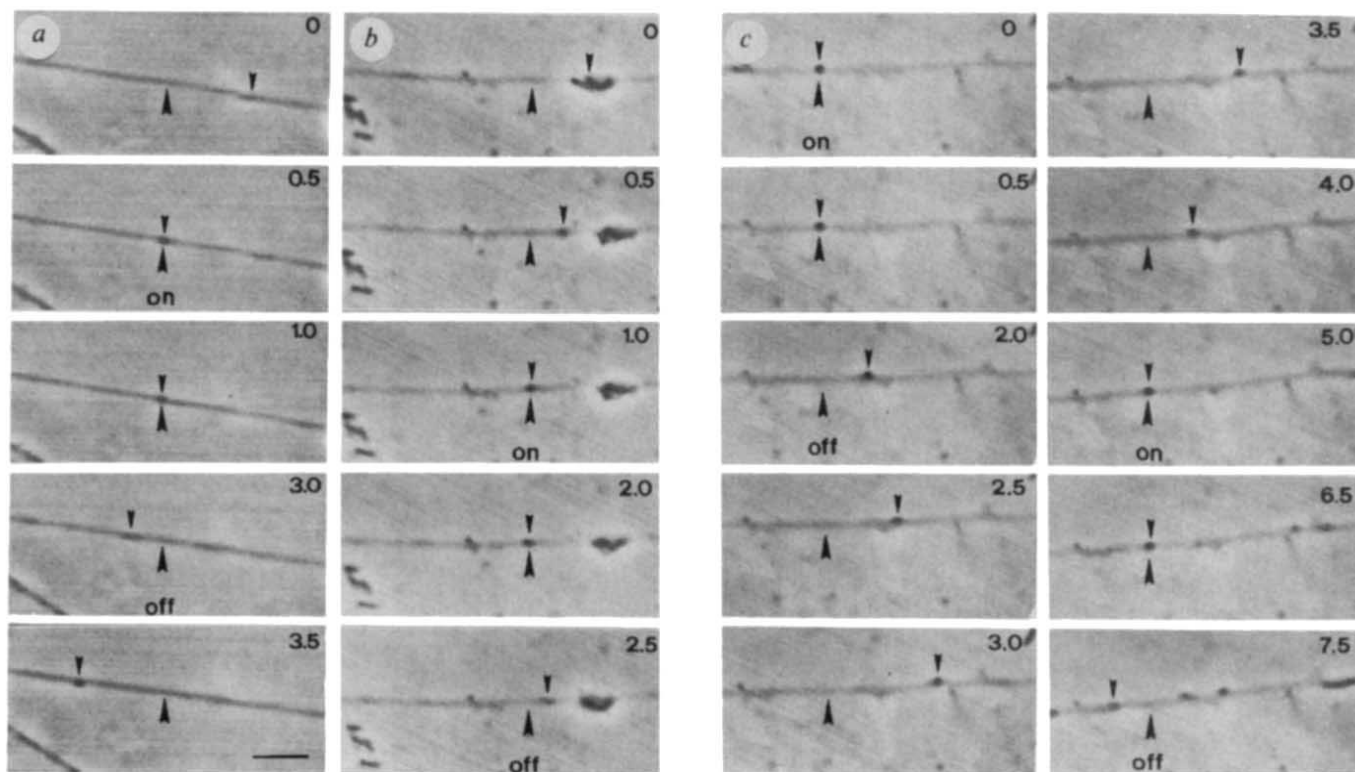


FIG. 2 Laser trapping of mitochondria in thin *Reticulomyxa* strands. Small arrowhead denotes the organelle; large arrowhead marks the position of the trap. Times in seconds are shown in the upper right-hand corner. 'On' and 'off' indicate the state of the laser trap. Scale bar, 5 μm . *a*, A mitochondrion travelling from right to left is trapped (frame 2), held in the trap for ~ 2 s, and moves on in the same direction after the trap is turned off. Its speed is $12.5 \mu\text{m s}^{-1}$ before and $10.7 \mu\text{m s}^{-1}$ after trapping. *b*, A mitochondrion reversing direction after being trapped for ~ 1 s (frames 3 and 4). Its speed is $8.7 \mu\text{m s}^{-1}$ before and $9.1 \mu\text{m s}^{-1}$ after trapping. *c*, This mitochondrion entered the trap from the left (frame 1), was released after ~ 1.5 s

and continued to move in the same direction (time 2–3 s). It then reversed its direction (between 3 and 5 s), and was trapped again at the same power (time 5–6.5 s).

METHODS. *Reticulomyxa* was grown and prepared for light microscopy as described³⁴. The laser trap system was the same as before^{17,18}, except that phase contrast optics were used. All experiments were recorded on video tape and analysed frame-by-frame. Sequences were photographed off the monitor using a Pentax 35 mm camera equipped with an automatic motor drive.

of 2–3 (combining errors in particle size distribution, force calibration, and power meter accuracy).

Our *in vivo* force measurement of $\sim 2.6 \times 10^{-7}$ dynes for a single (presumably dynein-like) motor is similar to, but more closely defined than, *in vitro* measurements for ciliary dynein of 1×10^{-7} dynes²⁵ and $0.5\text{--}10 \times 10^{-7}$ dynes²⁶, and indirect estimates of $2\text{--}9 \times 10^{-7}$ dynes^{27,28}. For kinesin motors driving microtubules across a glass surface, a force $> 6 \times 10^{-9}$ dynes was estimated¹⁴, whereas Block *et al.*²⁹, using kinesin-covered beads and sea-urchin axonemes, report a force of $> (1\text{--}50) \times 10^{-9}$ dynes. Kishino and Yanagida³⁰ estimate a force of $> 2 \times 10^{-8}$ dynes for a single myosin head. This compares with a force per head of $\sim 10^{-7}$ dynes for myosin in muscle during isometric contraction³¹.

Inherent in our *in vivo* technique is the potential to see the effects on motility due to differences in viscosity and in the number of motors in mitochondria. Estimates of viscosity or, more strictly, viscous resistance¹⁸ were made by dragging trapped mitochondria back and forth within strands. Values ranged from about 1,000 in stiff strands, to twice the viscous drag of water in loose amoeba droplets.

A striking feature of the data in Table 1 is the uniformity of the velocity (or stepping rate of the molecular motors) of mitochondria over many samples with different viscosities and also presumably with different numbers of motors. This constancy of speed at $\sim 10 \mu\text{m s}^{-1}$ indicates that individual active motors can run at narrowly fixed rates against a range of resistive loads (due to varying viscous drag for example). Similar behaviour was observed with kinesin driving microtubules *in vitro*¹⁴. This general behaviour is reminiscent of electrical synchronous motors. We expect such a rate-controlled motor to produce an increasing force as the load increases up to a point of overload, beyond which its behaviour should become erratic.

A Stoke's law calculation shows that mitochondria with one motor having a force of 2.6×10^{-7} dynes can run without overloading at a full velocity of $10 \mu\text{m s}^{-1}$ for all viscosities up to ~ 90 times that of water. Mitochondria with four motors can operate at $10 \mu\text{m s}^{-1}$ up to viscosities of ~ 360 times that of water without overload. These values seem reasonable in light of the approximate range of viscous resistances found in different amoeba samples.

There is qualitative evidence supporting the overload picture from the experiments measuring the ability of the trap to stop active mitochondria at gradually decreasing powers. Trapping ability is gauged from the approximate time it takes to capture a particle. We found that in more viscous strands that the range of possible stopping powers was restricted to high values, from ~ 220 to ~ 150 mW. Only in less viscous samples could trapping proceed down to values as low as ~ 55 mW. This is an indication that at high viscosity, mitochondria with more motors (3 and 4) are fully active and those with fewer (1 and 2) are in overload. The particles in overload are presumed to be those particles in viscous strands that are seen to move erratically at velocities markedly lower than $10 \mu\text{m s}^{-1}$. The data also suggest that at low viscosity, when all particles should be active, we are observing the trapping of particles with 1 or 2 motors at powers as low as 55 mW. This is reasonably consistent with our previously determined force for a single molecular motor of ~ 63 mW. In future more detailed *in vivo* studies of the effects of the numbers of motors and viscosity on the motility of individual organelles will probably make use of some type of more rapid automated force-measuring scheme. □

7. Brady, S. T., Pfister, K. K. & Bloom, G. S. *Proc. natn Acad. Sci. U.S.A.* **87**, 1061–1065 (1990).
8. Paschal, B. M., Shpetner, H. S. & Vallee, R. B. *J. Cell Biol.* **105**, 1273–1282 (1987).
9. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181–183 (1988).
10. Lye, R. J., Porter, M. E., Scholey, J. M. & McIntosh, J. R. *Cell* **51**, 309–318 (1988).
11. Euteneuer, U., Koonce, M. P., Pfister, K. K. & Schliwa, M. *Nature* **332**, 176–178 (1988).
12. Shroer, T. A., Steuer, E. & Sheetz, M. P. *Cell* **56**, 937–946 (1989).
13. Kishino, A. & Yanagida, T. *Nature* **334**, 74–76 (1988).
14. Howard, J., Hudspeth, A. J. & Vale, R. D. *Nature* **342**, 154–158 (1989).
15. Euteneuer, U. *et al.* in *Cell Movement* Vol. 2 (eds Warner, F. D. & McIntosh, J. R.) 155–167 (Liss, New York, 1989).
16. Ashkin, A. & Dziedzic, J. M. *Science* **235**, 1517–1520 (1987).
17. Ashkin, A. & Dziedzic, J. M. & Yamane, T. *Nature* **330**, 769–771 (1987).
18. Ashkin, A. & Dziedzic, J. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7914–7918 (1989).
19. Block, S. M., Blair, D. F. & Berg, H. C. *Nature* **338**, 514–517 (1989).
20. Koonce, M. P. & Schliwa, M. *J. Cell Biol.* **103**, 605–612 (1986).
21. Euteneuer, U., Johnson, K. B. & Schliwa, M. *Eur. J. Cell Biol.* **50**, 34–40 (1989).
22. Schliwa, M., Shimizu, T., Vale, R. D. & Euteneuer, U. *J. Cell Biol.* (in the press).
23. Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. & Chu, S. *Optics Lett.* **11**, 288–290 (1986).
24. Koonce, M. P. & Schliwa, M. *J. Cell Biol.* **100**, 322–326 (1985).
25. Kamimura, S. & Takahashi, K. *Nature* **293**, 566–568 (1981).
26. Oiwa, K. & Takahashi, K. *Cell Struct. Funct.* **13**, 193–205 (1988).
27. Hiramoto, Y. in *Cilia and Flagella* (ed. Sleight, M. A.) 177–196 (Academic, London, 1974).
28. Brokaw, C. J. in *Molecules and Cell Movement* (eds Inoue, S. & Stephens, R. E.) 165–179 (Raven, New York, 1975).
29. Block, S., Goldstein, L. S. B. & Schnapp, B. J. *J. Cell Biol.* **109**, 81a (1989).
30. Kishino, A. & Yanagida, T. *Nature* **334**, 74–76 (1988).
31. Oosawa, F. *Biorheology* **14**, 11–19 (1977).
32. Schliwa, M. & van Blerkom, J. *J. Cell Biol.* **90**, 222–235 (1981).
33. Euteneuer, U., Haimo, L. T. & Schliwa, M. *Eur. J. Cell Biol.* **49**, 373–376 (1989).
34. Koonce, M. P., Euteneuer, V., McDonald, K. L., Menzel, D. & Schliwa, M. *Cell Motil. Cytoskel.* **6**, 521–533 (1986).

ACKNOWLEDGEMENTS. We thank R. Wiegand-Steubing for collaboration in preliminary experiments. Supported by a grant from NSF to M.S. K.S. is a Feodor-Lynen Fellow of the Alexander von Humboldt Foundation.

Bead movement by single kinesin molecules studied with optical tweezers

Steven M. Block^{*†}, Lawrence S. B. Goldstein[†] & Bruce J. Schnapp^{‡§}

^{*} Rowland Institute for Science, 100 Cambridge Parkway, Cambridge, Massachusetts 02142, USA

[†] Department of Cellular and Developmental Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

[‡] Department of Physiology, Boston University School of Medicine, 80 East Concord Street, Boston, Massachusetts 02192, USA

KINESIN, a mechanoenzyme that couples ATP hydrolysis to movement along microtubules, is thought to power vesicle transport and other forms of microtubule-based motility^{1–6}. Here, microscopic silica beads⁷ were precoated with carrier protein^{8,9}, exposed to low concentrations of kinesin, and individually manipulated with a single-beam gradient-force optical particle trap^{10–12} ('optical tweezers') directly onto microtubules. Optical tweezers greatly improved the efficiency of the bead assay, particularly at the lowest kinesin concentrations (corresponding to ~ 1 molecule per bead). Beads incubated with excess kinesin moved smoothly along a microtubule for many micrometres, but beads carrying from 0.17–3 kinesin molecules per bead, moved, on average, only about $1.4 \mu\text{m}$ and then spontaneously released from the microtubule. Application of the optical trap directly behind such moving beads often pulled them off the microtubule and back into the centre of the trap. This did not occur when a bead was bound by an AMP.PNP-induced rigor linkage, or when beads were propelled by several kinesin molecules. Our results are consistent with a model in which kinesin detaches briefly from the microtubule during a part of each mechanochemical cycle, rather than a model in which kinesin remains bound at all times.

Silica beads, $0.2 \mu\text{m}$ in diameter⁷, were precoated by incubation with one or more proteins (bovine serum albumin or a mixture of casein and cytochrome c), which improves the

Received 23 April; accepted 2 October 1990.

1. Warner, F. D. & McIntosh, J. R. (eds) in *Cell Movement* Vol. 2 (Liss, New York, 1989).
2. Brady, S. T. *Nature* **317**, 73–75 (1985).
3. Vale, R. D., Reese, T. S. & Sheetz, M. P. *Cell* **42**, 39–50 (1985).
4. Scholey, J. M., Porter, M. E., Grissom, P. M. & McIntosh, J. R. *Nature* **318**, 483–486 (1985).
5. Shroer, T. A., Schnapp, B. J., Reese, T. S. & Sheetz, M. P. *J. Cell Biol.* **107**, 1785–1792 (1988).
6. Pfister, K. K., Wagner, M. C., Stenoi, D. L., Brady, S. T. & Bloom, G. S. *J. Cell Biol.* **108**, 1453–1463 (1988).

§ To whom correspondence should be addressed.

efficiency of moving actin filament⁸ or microtubule assays⁹. Precoating with these proteins permits movement at concentrations of kinesin that are 1–3 orders of magnitude below the thresholds for movement in their absence. The procedure could work by reducing surface denaturation of motor molecules, or the surface coating could facilitate binding of motors with a geometry more favourable to force production. Beads were incubated with kinesin purified from squid optic lobe and placed in suspension over a coverglass surface to which sea urchin axonemes or single taxol-stabilized squid microtubules had previously been bound (Fig. 1).

A diffusing bead was captured with the optical tweezers, which markedly reduced its brownian excursions, moved up to an axoneme or single microtubule and placed against it for about a second. Often, the bead would attach and begin moving spontaneously. After about a second, or as soon as movement was ascertained, the light from the laser was blocked by a shutter, whereupon, beads that had failed to bind generally diffused away. Beads that failed to bind to the microtubule or that ran off one end of a microtubule, or that were released before reaching the end, could be recaptured and re-tested. When beads carried small numbers of kinesin molecules, movement arising from random collisions of beads and microtubules was negligible.

Following bead placement on a microtubule we distinguished several outcomes: (1) the bead failed to bind, even after 2–5 trials; (2) the bead bound for a period of up to several seconds, but failed to move, or moved a distance too short to score

(<300 nm), and then was released; (3) the bead bound and moved, but was released from the microtubule before arriving at the end; and (4) the bead bound and moved to the end of the microtubule. Fewer than 4% of beads bound, moved and stopped moving, that is became 'stuck', or became attached to the coverglass next to the microtubule. These beads were not analysed further.

The relative frequency of the various outcomes depends strongly on the number of kinesin molecules carried by a bead (Fig. 2a). This number was computed by dividing the concentration of beads by the concentration of kinesin, a procedure that assumes that all kinesin molecules were in active form and bound to the silica beads. When incubated with relatively high levels of kinesin, corresponding to an average of ~17 kinesin molecules per bead, nearly all beads moved to the end of the microtubule or axoneme, a distance usually of >4 μm (depending on the point of deposition). As the mean number of kinesin molecules per bead was reduced, the fraction of moving beads that released early increased monotonically. At the lowest density (0.17 kinesin molecules per bead), all moving beads released. Also, as the mean number of kinesin molecules per bead was reduced, the fraction of beads that failed to bind increased monotonically.

Several lines of evidence suggest that motion at low density is due to single kinesin molecules. First, assuming that all kinesin added was bound to the beads after incubation and attached with a geometry favourable for interaction (which is expected to overestimate the amount of functional protein), movement

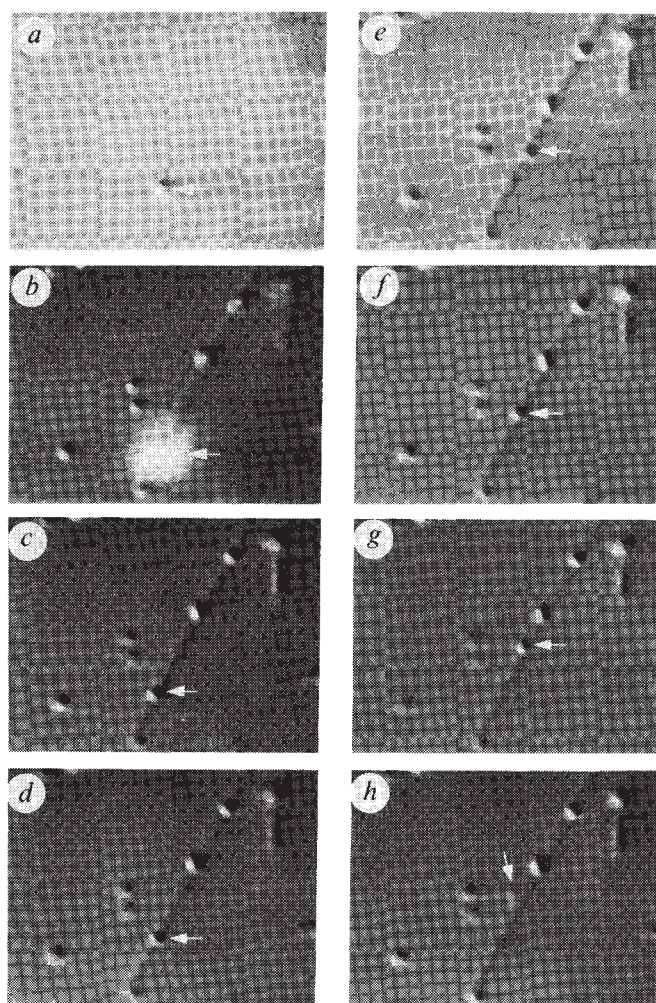


FIG. 1 Video sequence illustrating the use of optical tweezers to place a bead incubated with a low concentration of kinesin on an axoneme. *a*, The laser shutter was opened and a bead (indicated by the arrow) captured out of suspension; brownian movement stopped (frame 1, frames are 1 per 30 s). *b*, The microscope objective was focused near the coverglass and the x-y controls of the trap used to bring the bead over the axoneme; the reflection of laser light at the coverglass–water interface partially obscures the image (frame 41). The laser was shuttered, releasing the bead. *c–d*, The bead bound and began moving along the axoneme (frames 93, 157). Between *e* and *f*, the bead briefly dissociated from the microtubule, diffused, then bound again ~1 μm behind its earlier position (frames 242, 245). *g*, The bead continued forward along the axoneme (frame 334). *h*, The bead detached from the axoneme and diffused away (frame 341).

METHODS. Silica beads were preincubated for at least 1 min in BRB80 buffer (80 mM PIPES, 1 mM MgCl_2 , 1 mM EGTA; pH 6.8) supplemented with 2 mM ATP, 50 mM KCl, 0.5 mM dithiothreitol, and a blocking protein, comprising either 10 mg ml^{-1} BSA or 50 $\mu\text{g ml}^{-1}$ casein plus 50 $\mu\text{g ml}^{-1}$ cytochrome *c*. Beads were then incubated for 2 min with varying levels of kinesin (final dilution from stock, 1:100 to 1:5,000) and blocking protein. All containers, except for the coverglasses, were made from polypropylene. Kinesin was purified from squid optic lobe by microtubule affinity followed by sedimentation²³. Kinesin-coated beads were added to demembrated sea urchin axonemes²⁴, or to taxol-stabilized microtubules (10 $\mu\text{g ml}^{-1}$ taxol) free of microtubule-associated proteins and isolated from squid optic lobe²⁵ that had been adsorbed onto a coverglass. Silica beads were used at a final concentration of ~0.16 μm^{-3} , determined by counting beads from microscope fields. Concentration of kinesin heavy chain was determined from its band intensity and a standard curve obtained by video gel densitometry (Image-1, Universal Imaging) using a serial dilution of BSA as the standard. The number of kinesin molecules was calculated from this concentration by assuming a relative molecular mass of 110,000 for each heavy chain^{1,26–28}, and a stoichiometry of two heavy chains per molecule, based on molecular weight determination by analytical equilibrium sedimentation (refs 26 and 27 and our unpublished data). Specimens were observed by video-enhanced differential interference contrast microscopy²⁹, except that an extended-red Ultracon camera (series 70, Dage-MTI) was used, to pick up the reflection of the infrared laser. The design of the optical tweezers³⁰ is based on a Nd:YAG laser (wavelength 1,064 nm, Model C-95 YAGMAX, CVI Lasers). The diameter of the trapping zone was close to 1 μm ; it could be positioned anywhere in the field of view. Light intensity was varied with neutral density filters; the trapping was turned on and off by a millisecond shutter (Uniblitz, Vincent Associates). The z-position of the trap was adjusted to place the trapping zone in the microscope specimen plane. Trapping force was calibrated against Stokes' drag, in a manner similar to that described in ref. 12.

was observed at nominal dilutions corresponding to fewer than one kinesin molecule per bead, on average. Second, if releases occurred only in those beads that were moved by a single kinesin, one might expect the distance travelled before release to be independent of kinesin number, in conditions of low density. This is expected because the distance travelled reflects the properties of the kinesin that happens to be driving the movement, even though other kinesin molecules might be present elsewhere on the bead surface. This distance is $\sim 1.4 \mu\text{m}$, is independent of kinesin density, and the distribution of distances is roughly exponential, suggesting that beads have a constant probability of release per unit of travel (Fig. 2b). Third, the number of kinesin molecules on a bead should have a Poisson distribution (see Fig. 2c).

In an attempt to measure the force produced by kinesin molecules, we used the optical tweezers to try to stop moving beads at low kinesin densities. Quite often, beads moved right through the optical trapping zone shortly after placement on a microtubule, even when the laser was not quickly shuttered (see

Fig. 1). Sometimes, when the optical trap was positioned just behind a moving bead and the shutter opened, the bead would be pulled away from the microtubule and move rapidly (within one video frame) back into the centre position of the trap (Fig. 3). The bead could rebind to the microtubule and move again, as verified by releasing the trap and allowing the bead to continue. The trapping force, estimated at $0.5\text{--}5 \text{ pN}$ for these experiments, never stopped movement altogether: beads either continued right through the trapping zone at nearly constant speed or were pulled off. The mean distance travelled by a bead before being pulled off by this mechanism was $0.8 \pm 0.3 \mu\text{m}$ ($n = 27$), indicating that the application of the trap was hastening release from the surface, not merely catching beads that had spontaneously detached. This behaviour cannot be explained by the trap pulling the kinesin off the silica bead surface, because when the trap was positioned just behind kinesin-coated beads that were bound to microtubules in rigor, induced by the addition of the nonhydrolysable ATP analogue AMP.PNP, the same trapping force was insufficient to dislodge them, even at the

FIG. 2 a, Relative percentages of possible outcomes following placement of a bead on a microtubule as a function of kinesin density, shown as stacked bars: white, beads that failed to bind; light grey, beads that bound to a microtubule and moved less than 300 nm or not at all; dark grey, beads that bound, moved, then released before reaching the end of a microtubule; black, beads that bound to a microtubule and moved all the way to its end. In the experiment shown, 353 beads were tested; 339 were retained for analysis (see text). Of the latter, 212 moved on axonemes and 141 moved on single microtubules. At each kinesin density, number of beads tested was: 0.2, 114; 0.8, 61; 1.7, 58; 3.3, 74; 16.7, 32. b, Distribution of distances travelled before release from a microtubule. Data were pooled for all densities. Solid line, exponential fit to bins 2–9 of the histogram; length constant $1.1 \mu\text{m}$. Inset, semilog plot of average distances moved before release at each kinesin density; error bars, s.e.m. Global mean distance ($1.4 \pm 0.1 \mu\text{m}$; mean $\pm$ s.e.m., $n = 93$) indicated by the horizontal line. c, Comparison of possible outcomes with Poisson statistics. Upper panel, fraction of beads that failed to bind ($\circ$) compared with the probability that the bead carried no kinesin (---); fraction of beads that moved to the end of a microtubule ($\blacksquare$) compared with the probability that a bead carried two or more kinesins (.....). Lower panel, fraction of beads that released from a microtubule after movement ($\bullet$) compared with the probability that a bead carried one kinesin (—). Kinesin molecules are assumed to function independently and be attached at random points on the bead surface. The probability that a bead carried no kinesin is given by $P(0, x) = e^{-x}$, where x is the average number of kinesins per bead and $P(n, x)$ is the Poisson distribution. The probability that a bead carried one kinesin is $P(1, x) = xe^{-x}$; for two or more, it is $(1 - e^{-x} - xe^{-x})$. The model involves no free parameters that may be adjusted to fit. Note that at the lowest density tested, the chance that a bead carried no kinesin at all is $\exp(-0.17) = 0.85$, while the chance that it carried just one is $(0.17) \exp(-0.17) = 0.14$. This compares favourably with the observed fraction of beads that failed to bind to a microtubule, 0.87, and the fraction that moved and released, 0.09. At this density, the chance that two or more motors might be on one bead is 0.01. The tendency for the data in Fig. 2c to lie to the right of the curves can be explained by assuming that a fraction of the added kinesin was inactive or adhered to the walls of the container(s), or was bound to beads on a part of their surface distant from the point of contact with the microtubule. This would mean that a nominal density of, say, three kinesin molecules per bead actually gave one operational kinesin molecule per bead. The data might equally well be fit by assuming that the kinesin densities are correct as given, but that beads carrying three or so kinesins (as opposed to one) will release. In fact, given the large surface area of a bead ($1.3 \times 10^5 \text{ nm}^2$), it is improbable that two kinesin heads will find themselves within a few nanometres of one another, even at substantially higher kinesin densities. We have also observed movement of still larger beads ($0.4 \mu\text{m}$ diameter) carrying kinesin at nominal densities corresponding to only a few molecules per bead.

METHODS. Axonemes and single microtubules were both used in the same microscopic preparation to compare the two substrates with a single population of beads. Distributions of outcomes and the mean distance travelled before release were not statistically different for the two substrates ($P < 0.05$; t -test), so data were pooled for the graphs shown. Trials with beads that were placed too close to the end of a microtubule or axoneme (that is, within $4.5 \mu\text{m}$, equivalent to three length constants) and subsequently moved to the end were not scored.

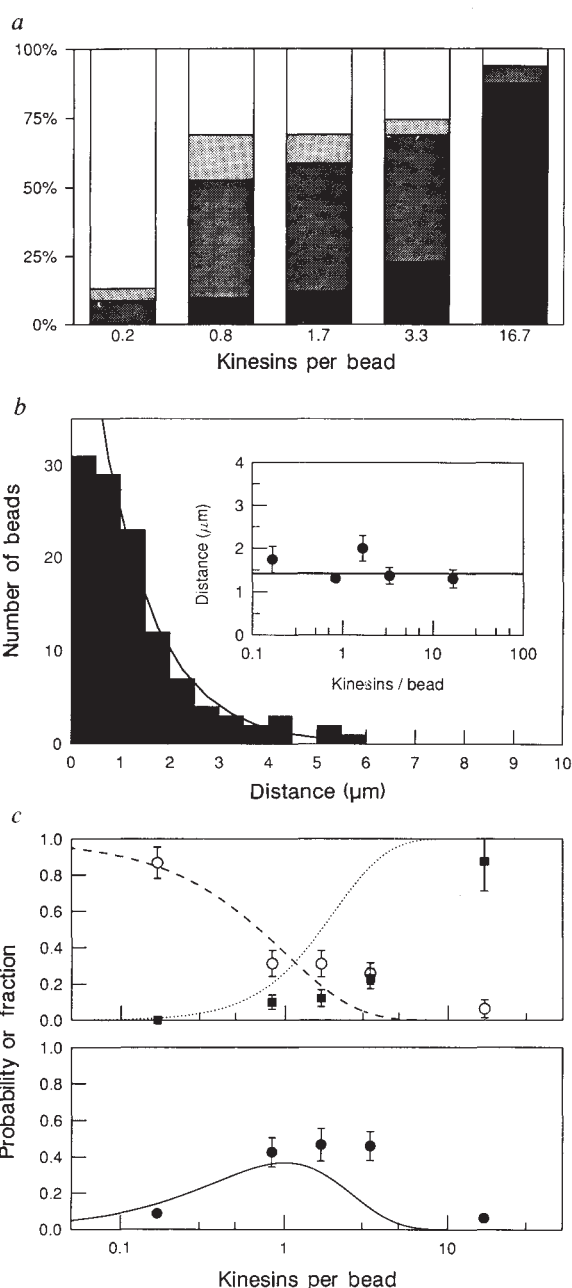
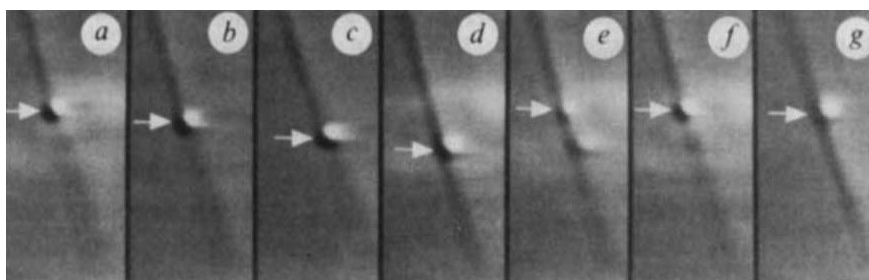


FIG. 3 Video sequence showing a bead being pulled off the axoneme and back into the centre position of the optical trap. *a*, Bead (arrowed) is deposited by optical tweezers on an axoneme. Some of the laser light can be seen in the region of the bead as a result of scattering from the coverglass (frame 1). *b–c*, The trap is shuttered and the bead begins to move (frames 19, 62). *d*, The shutter is opened with the trap in its original location (frame 78). *e*, Bead is pulled back into the centre of the trap within one frame interval; two bead images arise from the video interface (frame 79). *f*, Bead rebinds to the axoneme (frame 80). *g*, Trap is shuttered and bead continues movement (frame 120). The shape of the potential well experienced by the bead is such that maximum trapping forces



are produced in regions near the outer margins of the light spot; the trapping force is zero at the centre of the well.

lowest kinesin density.

There are several possible reasons for beads diffusing away from a microtubule after moving on average only 1.4 μm . One is that defects in the microtubule lattice or the presence of binding proteins on the microtubule leads to release of kinesin from the microtubule. This seems unlikely in view of the fact that the mean distance travelled along both axonemes and microtubules was the same (Fig. 2), and because movements over much longer distances have been observed for moving microtubule assays driven by single kinesins (our unpublished data, and ref. 9). A second possibility is that the kinesin might have detached from the silica surface rather than from the microtubule. This seems unlikely in view of the observation that a rigor linkage could not be broken. Third, kinesins might function for a fixed number of cycles before becoming inactivated. This is a possibility that we cannot exclude with our data but do not favour, as it is difficult to reconcile with the kinesin density-dependence of movement (Fig. 2*a*). Finally, the kinesin molecule might detach briefly from the substrate during each mechanochemical cycle. In this model (the stroke-release model) the bead will occasionally diffuse far enough away from the microtubule to preclude attachment for the next power stroke. (This model does not specify whether the kinesin heads work cooperatively or independently.) Alternatively, the kinesin might remain bound during each mechanochemical cycle, but in a loosely coupled state whose binding energy is comparable with thermal energy, kT : this amounts to a semantic distinction for the present discussion.

Kinesin's ability to move tiny cytoplasmic organelles, presumably too small to accommodate many motors¹³, the tiny nanometre-sized side-to-side excursions measured for moving beads¹⁴, and the demonstration that microtubules can be driven for long distances by single kinesin molecules, suggest that the kinesin heads might walk 'hand over hand', interacting cooperatively in such a way that the substrate is never released^{9,15}. This is in contrast to the muscle force-generating protein myosin, whose heads are detached during a large fraction of its mechanochemical cycle^{16,17}. A stroke-release model for kinesin is, however, consistent with all the available data, provided head detachment is sufficiently brief. Conversely, it is difficult to reconcile spontaneous and trap-induced releases with 'hand-over-hand' models lacking a detached state (see Fig. 4*a*). Assume kinesin were to step discretely. If one estimates the step distance at ~ 10 nm, a value chosen to be comparable to the kinesin head length¹⁸ and to one estimate of the myosin step size¹⁹, then a mean distance to release of 1.4 μm would correspond to a release probability of roughly 1% per stroke (10 nm per 1.4 μm), and the distance would be distributed exponentially. Monte Carlo simulations of a sphere diffusing near a long cylinder suggest that a bead 0.2 μm in diameter could detach

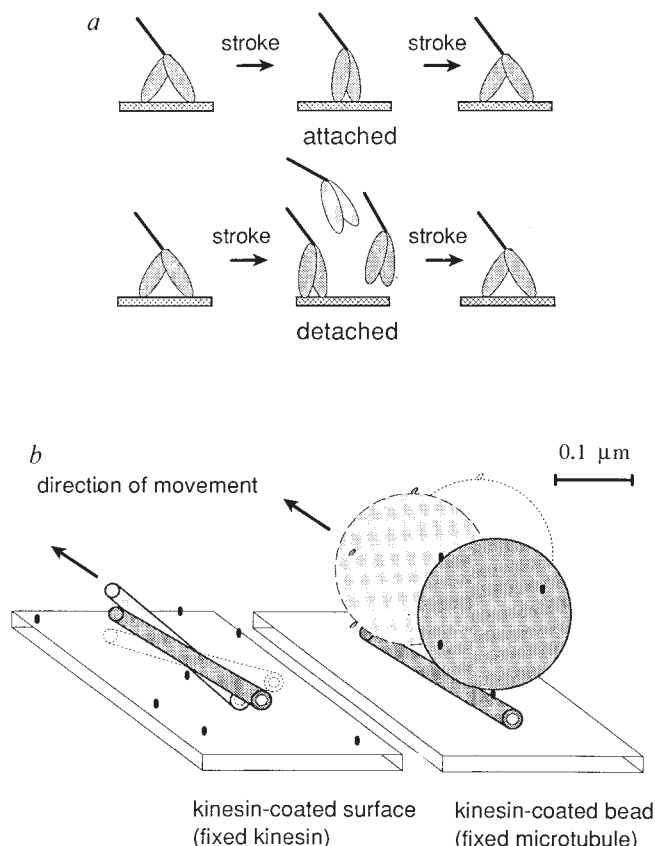


FIG. 4 *a*, Differences between two classes of models for kinesin movement. In 'hand-over-hand' models (upper panel), at least one domain of the kinesin molecule remains attached at all times throughout the cycle. While two heads are shown acting cooperatively, single-head models incorporating two (or more) tubulin binding sites are equally plausible. In stroke-release models (lower panel), the molecule becomes completely detached from the substrate during a brief part of each cycle, during which time it can diffuse in its environment. The drawing is not intended to give a detailed representation of any one particular (or complete) sequence of mechanochemical events. *b*, Cartoon illustrating the comparative effects of diffusion on moving microtubule and moving bead assays at low densities of kinesin. Beads, microtubules and kinesin heads are all drawn to relative scale, except for lengths of the microtubules, which have been foreshortened for the purpose of illustration. Images are shown at successive instants, with displacements exaggerated. A moving microtubule (left) is seen to pivot about an axis through its point of interaction with a kinesin head (nodal point pivoting). Release from the kinesin, followed by rotational diffusion of the microtubule about its long axis, will still present the tubulin surface for later binding. Rotation about this axis is expected to be relatively rapid as a result of the small diameter of the microtubule. A moving bead (right) is also subject to rotational diffusion, which carries the kinesin head away from the microtubule into a position that prevents subsequent binding. Because translational diffusion coefficients depend most strongly on the longest linear dimension³¹, a microtubule 2 μm long has a smaller translational diffusion coefficient normal to its long axis than a bead of 0.2 μm diameter by nearly a factor of 10 (ref. 32). This will lead to a reduced tendency for the microtubule, compared with the bead, to leave the surface.

from tubulin and diffuse for $\sim 1 \mu\text{s}$, yet still have a 99% chance of returning to the substrate within 10 ms (our unpublished data), a time sufficiently short (compared with the ATP turnover rate) for movement to appear largely unaffected. Qualitatively, the occasionally jittery character of bead motion driven by low numbers of kinesin molecules, including events of the kind shown in Fig. 1e–f, might be explained by a small number of longer-lived detached states. Detailed comparisons await data from tracking experiments that can follow fine structure of the motion¹⁴. Until now, it has not been possible to relate this motion to the action of single molecules.

The spontaneous bead detachments that we observe might be due to kinesin molecules operating with just a single active head, if, for example, the second head were inactivated by being bound to the bead surface. In this fashion, cooperativity between the two heads would be precluded and single-headed interactions would lead to premature release. Implicit in this interpretation is that the two kinesin heads are capable of independent motion, as seems to be the case for myosin⁸. This suggestion could be tested by studying motion driven by genetically engineered, single-headed kinesin²⁰.

How can movement over relatively long distances (5 μm and more) in moving microtubule assays driven by single molecules be reconciled with our observations of spontaneous release after $\sim 1.4 \mu\text{m}$? Figure 4b illustrates two effects that conspire to keep a microtubule moving near the glass, thereby obscuring the releases revealed by our moving bead assays. First, despite its narrow diameter, a microtubule has a smaller translational diffusion coefficient along the direction normal to the surface than a bead. Second, diffusional rotation of a moving microtubule about its long axis will present an equivalent surface lattice of potential binding sites, whereas rotation of a bead carrying a few kinesin molecules can destroy the possibility of favourable binding. Howard *et al.*⁹ recognized that movement driven by single kinesins might be consistent with a stroke-release model, and suggested that a detachment of less than 1 ms could be accommodated. Our data favour precisely such a model, and are fitted well by a microsecond-long inactivation of kinesin followed by a rebinding step that is diffusion-limited, taking a variable time, up to milliseconds. It is not clear how release time relates to measured rate constants of the biochemical cycle^{21,22}. We found no significant dependence of the speed of bead movement on the kinesin density; average speeds all fell in the range 0.35–0.45 $\mu\text{m s}^{-1}$. The rate-limiting step for movement seems not to be associated with the reassociation step following release.

The existence of a detached state calls into question the feasibility of using (constant) tension probes, such as optical tweezers or glass microneedles, to measure the force of translocation for beads or microtubules that are driven by small numbers of motors. This is because beads or microtubules are likely to be driven backwards once molecules detach. Measurements of the average force generated by larger numbers of kinesin molecules may ultimately be more meaningful. □

Received 1 July; accepted 12 September 1990.

- Vale, R. D., Reese, T. S. & Sheetz, M. P. *Cell* **42**, 38–50 (1985).
- Vale, R. D. *et al.*, *Cell* **43**, 623–632 (1985).
- Vale, R. D. *et al.*, *Rev. Cell Biol.* **3**, 347–378 (1987).
- Shroer, T. A. *et al.*, *J. Cell Biol.* **107**, 1785–1792 (1988).
- Scholey, J. M. *Nature* **318**, 483–486 (1985).
- Brady, S. T., Pfister, K. K. & Bloom, G. S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1061–1065 (1990).
- Stöber, W., Fink, A. & Bohn, E. *J. Colloid Interface Sci.* **26**, 62–69 (1968).
- Toyoshima, Y. Y. *et al.*, *Nature* **328**, 536–539 (1987).
- Howard, J., Hudspeth, A. J. & Vale, R. D. *Nature* **342**, 154–158 (1989).
- Ashkin, A. & Dziedzic, J. M. *Science* **235**, 1517–1520 (1987).
- Ashkin, A., Dziedzic, J. M. & Yamane, T. *Nature* **330**, 769–771 (1987).
- Block, S. M., Blair, D. F. & Berg, H. C. *Nature* **338**, 514–518 (1989).
- Miller, R. H. & Lasek, R. J. *J. Cell Biol.* **101**, 2181–2193 (1985).
- Gelles, J., Schnapp, B. J. & Sheetz, M. P. *Nature* **331**, 450–453 (1988).
- Schnapp, B. J., Crise, B., Sheetz, M. P., Reese, T. S. & Khan, S. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Huxley, H. E. & Faruqi, A. R. *Rev. Biophys. Bioeng.* **12**, 381–417 (1983).
- Cooke, R. *C.R.C. Crit. Rev. Biochem.* **21**, 53–118 (1986).
- Hirokawa, N. *et al.*, *Cell* **56**, 876–878 (1989).

- Huxley, A. F. & Simmons, R. M. *Nature* **233**, 533–538 (1971).
- Yang, J. T. *et al.*, *Science* **249**, 42–47 (1990).
- Kuznetsov, S. A. & Gelfand, V. I. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8530–8534 (1986).
- Hackney, D. D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6314–6318 (1988).
- Schnapp, B. J. & Reese, T. S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1548–1552 (1989).
- Bell, C. W. *et al.*, *Meth. Enzym.* **85**, 450–474 (1982).
- Vale, R. D., Schnapp, B. J., Reese, T. S. & Sheetz, M. P. *Cell* **40**, 559–569 (1985).
- Kusnetsov, S. A. *et al.*, *EMBO J.* **7**, 353–356 (1988).
- Bloom, G. S. *et al.*, *Biochemistry* **27**, 3409–3416 (1988).
- Kosik, K. S. *et al.*, *J. Biol. Chem.* **265**, 3278–3283 (1990).
- Schnapp, B. J. *Meth. Enzym.* **134**, 561–573 (1986).
- Block, S. M. in *Noninvasive Techniques in Cell Biology*, 375–401 (Wiley-Liss, New York, 1990).
- Berg, H. C. in *Random Walks in Biology*, Ch. 4 (Princeton University Press, 1983).
- Brennen, C. & Winet, H. A. *Rev. Fluid Mech.* **9**, 339–398 (1977).

ACKNOWLEDGEMENTS. We thank H. Berg for silica beads, J. Scholey for sea urchin axonemes, T. Reese for squid optic lobes, R. Bechtold and C. Godek for technical assistance, and H. Berg and T. Reese for comments on the manuscript. S.M.B. is supported by the Rowland Institute for Science. B.J.S. and L.S.B.G. are supported by the National Institutes of Health.

CORRECTION

Light-emitting diodes based on conjugated polymers

J. H. Burroughes, D. D. C. Bradley, A. R. Brown, R. N. Marks, K. Mackay, R. H. Friend, P. L. Burn & A. B. Holmes

Nature **347**, 539–541 (1990)

IN the published version of this paper the name of one of the authors was incorrectly cited as P. L. Burns rather than P. L. Burn. His address is the University Chemical Laboratory, Cambridge, rather than the University Chemistry Laboratory as stated.

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully. Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Birth of the Gods

Fred Hoyle

The Cosmic Winter. By Victor Clube and Bill Napier. Blackwell: 1990. Pp. 307. £15.95, \$24.95.

THE Solar System possesses an outlying cloud of comets reaching to a distance of perhaps a light year from the Sun — most aphelion distances are probably less than half of that. Otherwise the comet cloud would long ago have been tidally sheared away from the Solar System in an approach to a galactic molecular cloud. But Clube and Napier think that the Solar System has indeed repeatedly lost and then recaptured its cometary cloud. Recapture not being a simple reversal of loss, I would rather think the main cloud to be smaller than assumed.

Comets with near-zero angular momentum possess orbits that plunge deep into the Solar System, where planetary perturbations cause most of them to be ejected. A small fraction evolve with decreasing aphelion distances. Their periodicities shorten greatly, they become more and more degassed, ultimately leaving inert residues thought to be about a kilometre in size. Such residues may be destroyed in planetary encounters or generate more debris through collision with bodies in the asteroidal belt.

This process is enacted in a few million years, and, given the age of the Solar System, the cometary cloud would have been denuded of comets of small angular momentum if a resupply were not generated by perturbations from outside. The tidal gravitational fields of molecular clouds are the most effective regenerating influence. Such fields depend on the position of the Sun with respect to the galactic disk and spiral arms as well as to the position of individual molecular clouds. Clube and Napier consider this generates periodicities in the supply of comets of small angular momentum, of which they identify a long period of about 230 million years associated with the rotation of the Galaxy, within which there is a shorter period of 15 million years that alternates between effects that are strong and weak.

These periodicities are ultimately reflected in the bombardment of the Earth

by cometary bolides, by cometary fragments and by cometary dust. To this Clube and Napier attribute a very wide range of terrestrial phenomena — mass biological extinctions, changes in the mode of biological evolution, ice ages, orogenic episodes, shifts in plate tectonics and geomagnetic reversals — thereby satisfying Tommy Gold's dictum that in

products. A careful analysis of L and D alloseucine might reveal something useful, but no analysis I have seen is of such a degree of sensitivity.

But spectacular as all this may be, it pales beside the main thrust of the book. First, the authors postulate "giant comets" which in mass are comparable to asteroids, say 10^{21} grams, rather than the lower value of 10^{18} – 10^{19} grams that would be assigned to big comets from observation. Second, they say that such a comet became deflected into an essentially irreversible orbit inside that of Jupiter about 6,000 years ago. Comet Encke and the Taurid meteor streams are regarded as the surviving remnants of the degeneration of such an object. Any surviving large piece

of debris, of which Comet Encke may be just one among many, follows a changing orbit that intersects the plane of the Earth's orbit in two nodes. The solar distance of one or other of the nodes may coincide almost precisely with the radius of the Earth's orbit for a duration of about a century, during which interval the Earth will be exposed to debris incident at high speed from without, a challenge to what the authors refer to as the present-day closed-box mentality of the human species. And if there are many such objects with a pair of dangerous nodes, impacts can continue over an extended time interval.

In June 1975 as many one-ton boulders hit the Moon in five days as had hit the Moon in the previous five years. The late June date of this event indicates

a connection with the Taurid stream. The Tunguska event occurring on the night of 30 June 1908 was surely also connected with the Taurid stream. Because of their low tensile strength, bolides of stone shatter into many pieces if incident on the Earth's atmosphere, at an altitude of about 10 kilometres, except for bolides of kilometre size which penetrate to ground level and produce craters. The heat released almost instantaneously into the atmosphere is that which would be obtained by burning a lump of coal with a mass about a hundred times greater than the bolide. The mass for the Tunguska event is thought to have been about 30 megatons. The resulting damage on the ground from heat and from the atmospheric blast made human survival unlikely to a distance of about 50 kilometres. Extensive fires must have been started in the surrounding forest. It



Zeus — Does the king of the Gods owe his origin to a particularly bright comet?

the advancing of new hypotheses it does not pay to think timidly. On the Cretaceous-Tertiary (K/T) boundary, Clube and Napier are scathing in their comments about the mistakes which have been made by media-sponsored bandwagons that have been set rolling in the past decade, and on the persistent failures of those involved to reference similar work on the subject before their own. Despite their care in these matters there is one important point on which I suspect the authors themselves have been deceived. On page 227 they remark that amino acids found in K/T boundary deposits are of nonbiological origin, by which I suppose they mean there are nonbiological amino acids and that the amino acids are racemic. This is a non-sequitur because fossilized biomaterial goes racemic in a time-scale of only 100,000 years, and nonbiological amino acids also appear in the degeneration

was an event to recall the ancient myth of Phaethon, who being entrusted to take charge of the chariot dragging the Sun across the sky drove too low and set the Earth on fire.

On 25 June AD 1178, five excellent men of Kent saw fire and sparks break out repeatedly from between the horns of a new moon. The body of the moon "writhed and throbbed like a wounded snake" wrote Gervase of Canterbury, a "phenomenon that was repeated a dozen times or more". After the flame had twisted randomly for a while the Moon returned to normal, except that the crescent was now dark instead of bright as it had formerly been. In modern times, satellites have detected a very fresh crater, named (ironically) Giordano Bruno, on the far side of the Moon, which seems to have been produced in an impact of the 100,000 megaton class, sufficient if it had hit the Earth to have devastated at least an entire continent. Clube and Napier do not mention that this event demonstrates the correctness of Gold's view of the dusty nature of the lunar regolith, for it was surely immense clouds of thrown-up dust reflecting sunlight that the men of Kent took to be the fire and the sparks. The basalt flows, so dear to the NASA establishment, would not have produced anything of the sort. It is interesting that an inexpensive but correctly motivated library search would have solved the problem of the nature of the lunar surface, whereas the immensely expensive NASA lunar missions did not — the difference one might say between an open-box mentality and a closed-box one.

Clube and Napier regard a date of about 1000 BC as being crucial in human history. Before 1000 BC, and especially in the third millennium BC, bolide impact was far more common than it has been in the period since. In earlier times the skies were ablaze with comets, the many companions of Comet Encke. The emergence of particle jets from comets was interpreted as being similar to the destructive effects of bolide impacts on the Earth. The comets were gods and in firing away with the evaporated jets they were at war with each other. This was the background to ancient myths of the war of the gods. One particular comet lasted in brilliance longer than the others and this became Zeus, king of the gods. On pages 198–199, the authors compare their views on the genesis of ancient myths with the usual explanations, and I must say the usual looks very thin when compared with the brew on offer. Were the bolts of Zeus nothing more than lightning flashes which have only a minuscule probability of killing anyone? Or were they Tunguska-like events about which even hard men would be pressed to maintain their composure. Such events causing forest fires in

regions where metallic ores were exposed at the surface could produce reduction to metals. Was this how metal smelting was discovered, in cosmically generated furnaces? Where, I ask myself, did the expression "bolt from the blue" come from? Is there a similar expression in other languages? Splendid stuff! The authors are to be congratulated on having produced a book which seems well calculated to annoy a commendably large number of people. □

Sir Fred Hoyle is at 102 Admiral's Walk, West Cliff, Bournemouth BH2 5HF, UK.

A passage from India

David Lindley

Chandra: A Biography Of S. Chandrasekhar By K. C. Wali. University of Chicago Press: 1990. Pp. 352. \$29.95, £23.95.

A NINETEEN-year-old prodigy, travelling by boat from India to England where he is to take up graduate studies in astrophysics, discovers a new, profound and surprising fact about the nature of stars. After a few years at Cambridge, as his reputation is beginning to spread, he makes his discovery known, and, to his great dismay, is attacked — mocked in fact — by a man who was supposed to be his mentor. And although the logic of the young man's argument is irrefutable, the physics of the matter is somewhat abstruse, and the position of his attacker so lofty, that he finds himself largely abandoned.

In this literate and engaging biography K. C. Wali tries, as a novelist would, to make of this episode in the life of Subramanyan Chandrasekhar a key to understanding his subsequent career. The story of Chandra's discovery — that stars above a certain critical mass could not settle down into quiet old age as white dwarfs, but would collapse further into some then inscrutable new state of matter — and of how Sir Arthur Eddington disparaged it not for reasons of physics but because he could not bring himself to believe that nature would allow such a catastrophe, is part of the folklore of astrophysics. Had Eddington, a great expert on relativity, not thrown his weight against the idea, black holes might have been recognized as an astrophysical reality twenty years earlier. But Chandra, though he never doubted his own rightness, chose to disengage from any bitter public dispute with Eddington, and embarked on no lifelong vendetta.

Instead, he moved in 1937 to the United States, first to Yerkes Observatory in Wisconsin and then to the University of

Chicago, where, for half a century, he has cultivated what might be called the high style of research in mathematical physics. He immerses himself in the study of some seemingly remote subject and, after some years of intellectual labour, delivers an encyclopaedic monograph: *Radiative transfer*; *Ellipsoidal figures of equilibrium*; *The mathematical theory of black holes*.

It may be, as Wali suggests, that Chandra's decision to avoid the frenzied and disputatious frontiers of science, finding his satisfaction instead in the full comprehension of each complex problem that he tackles, was a reaction to his fierce encounter with Eddington. But the habit of self-reliance seems to have been instilled at an early age. In British India, when Chandra was growing up, the best schools were reserved for children of the British and of local princes; Chandra was educated mostly at home, by a learned but stern father and then by private tutors. Even then, he supplied his own discipline, working on his exercises in mathematics beyond his father's requirements.

And later, when Chandra's reputation was established in Europe and the United States, he resisted efforts to draw him back to any science position in India, principally, it seems, because he wished to stay aloof from the incessant squabbling that afflicted Indian science. In the middle decades of this century there were a number of distinguished Indian scientists, including H. Bhabha, S. N. Bose, M. N. Saha and Chandra's uncle, the Nobel laureate C. V. Raman, most of whom seemed to spend a good deal of their time scheming against the others. Although Chandra at one time likened the quality of life in the United States to distilled water, his position in Chicago afforded him the calm to indulge his intellectual tastes.

Wali writes persuasively of the difficulties encountered by an Indian trying to make a career in western science. Chandra had to endure patronizing British officials who hinted that he would soon find himself out of his depth, and scientists too respectful of Eddington and their own caste system to come forthrightly to his defence. And when he had made his name in the United States he was sniped at by scientists and bureaucrats back in India for his refusal to return. Those who know Chandra through his scientific reputation will find here many clues to an apparently distant character. His pleasantly dictatorial stewardship of the *Astrophysical Journal*, which he edited for many years, shows a man who will do no favours and waive no rules. He offended many astronomers, who had until then regarded the journal as an outlet for whatever they deemed suitable, but he made the *Astrophysical Journal* into a serious publication. And if it seems

like arrogance for Chandra to have been so determined to do the right thing, we have to recognize that, almost always, he is right.

Wali's account of Chandrasekhar's life is a social and intellectual education; it recalls another era, and a scientific style that is out of fashion (no fifty *Physical Review Letters* a year from Chandra). The only failure of the book is that it presumes a certain familiarity with Chandra's achievements. Readers may well want to know what Chandra did after the momentous work on white dwarfs and the battle with Eddington, but Wali describes an eminent, greatly admired scientist without sufficiently explaining the eminence and admiration. It would be a shame if this kept a large audience from the book: there is much to be enjoyed in it. □

David Lindley is an Associate Editor of *Nature*.

■ Just published by Chicago University Press is volume 5 of the selected papers of S. Chandrasekhar: *Relativistic Astrophysics*. Price is £59.95, \$86.25 (hbk); £23.95, \$34.50 (pbk).

Astro aesthetics

David W. Hughes

Universe. By Michael Rowan-Robinson. Longman: 1990. Pp.180. £17.95. Published in the United States by Freeman as **Our Universe: An Armchair Guide**. \$24.95.

ASTRONOMERS usually, and quite understandably, wince when confronted with the following quotation.

*When I heard the learn'd astronomer,
When the proofs, the figures, were ranged
in columns before me,
When I was shown the charts and diagrams,
to add, divide and measure them,
When I sitting heard the astronomer where
he lectures with much applause in the
lecture room,
How soon unaccountable I became tired and
sick,
Till rising and gliding out I wander'd off by
myself,
In the mystical moist night-air, and from time
to time,
Look'd up in perfect silence at the stars.*

(W. Whitman, "By the Roadside".)

Rowan-Robinson uses this quotation to conclude the preface of *Universe*. The quotation is redolent with hints of C. P. Snow's two cultures. One culture is supposedly standing in awe and wonder and is receptive to beauty, and the other culture, replete with instruments, graph paper and computers, is only happy when it measures, dissects and understands. It is easy to see which side Whitman is on. He is liberal with his artistic sneers. To Whit-

man, the scientist is a deflowerer as well as a philistine. Whitman clearly knew too few. The joy of reading *Universe* was also not his.

Rowan-Robinson has taken great pains, not only to stride over the cultural divide, but also to pull the two sides together. His underlying thesis is that ignorance is not bliss; beauty is enhanced by understanding, and the more you know, the more the joy increases.

Universe is divided into twenty chapters, each devoted to a single object. Comets rub shoulders with the nearest star to the Sun, Vega with Mira, the ring nebula with the Crab, Andromeda with the Magellanic Clouds, radio galaxies with quasars and the Starburst galaxy with the Virgo cluster. The selection favours the rare, the exotic and the spectacular. New stars, demon stars, variability, explosions, gas jets and death throes are in, the mundane — the normal, slow, astrophysical drudgery of ordinary stars and planets — is out. In fact all the planets are out. Something as small, rocky and wet as the Earth is merely regarded as a punch bag for incoming asteroidal fragments and a supporting surface for astronomical observatories. Our life and the possibility of life elsewhere is not considered. The only thing that Rowan-Robinson seems to find really exciting within the nearest four light years is our once-in-a-lifetime

encounter with Halley's Comet. But this fact underlines one of the joys of the book. It is a personal selection and a personal romp through the visual heavenly wonders. I should stress the word 'visual'. The ideas that underlie the origin and evolution of the Universe and underlie our ever-changing perception of what the Universe contains, feature much less in this book than those contents of the Universe that can be easily imaged.

Each of the twenty objects is dealt with in nine pages. To say that the book is well illustrated is an understatement. It is a visual delight and colour abounds. Even Whitman would appreciate the astronomical aesthetic enhancement that has resulted from the employment of large telescopes and spacecraft instrumentation. Even he would realize that standing moistly and gazing silently is a poor



Shooting stars — Historical woodcut of the Leonids meteor shower, reproduced from *Universe*.

substitute for the more modern endeavours of the astronomical community.

The two cultures still exist. They still sneer and snipe at each other from their close-guarded battlements. But bridges can be built and astronomers have an easier task than the majority of scientists.

Rowan-Robinson has produced a beautiful, stimulating and rewarding book. It is a feast for both the mind and the eye. The scientist is treated to a host of literary quotations and artistic reproductions, and the liberal artist to a careful exposition of the physical and chemical mechanisms that underpin the activities of the exotic bodies in the Universe. There is something in this book for us all. □

David W. Hughes is Reader in Astronomy in the Department of Physics, University of Sheffield, S3 7RH, UK.

Past and future

Huw Price

The Arrow of Time. By Peter Coveney and Roger Highfield. Allen: 1990. Pp. 378. £14.95.

Time Journeys. By Paul Halpern. McGraw-Hill: 1990. Pp. 153. Hbk \$19.95, £19.95, pbk \$11.95, £10.95.

THE arrow of time is one of the big unclaimed prizes of modern physics. The problem is to reconcile the temporal asymmetry of thermodynamics with the apparent temporal symmetry of fundamental physical theories. Some major players have wrestled with the issue over the past century or so, but it is still up for grabs — and very much in the air of late, having been discussed in recent books by Stephen Hawking (*A Brief History of Time*) and Roger Penrose (*The Emperor's New Mind*), among others.

The Arrow of Time, subtitled "A voyage through science in search of time's greatest mystery", is a lively guide to the current state of play. Its authors are a scientist and a distinguished science journalist. The former, Peter Coveney, is a member of a team that has claimed a novel and fruitful approach to the game (and whose captain, Ilya Prigogine, provides a foreword to the book). The viewpoint is therefore partisan: the authors soon make it clear that they want to argue that the key to the puzzle lies in some of the insights of Prigogine's Brussels school, and in related work in the theory of chaos. This is a bad omen, for as we will see, someone who understands the rules of the game (that is, who grasps the real mystery about temporal asymmetry) should be able to see that the proposed solution is simply addressing the wrong issue. For all its intrinsic interest, it cannot provide an answer to the main question. The result is a bit like *Babe Ruth's Guide to Mysteries of Cricket*: fine on baseball, but not what we were promised.

The book's basic mistake is to fail properly to distinguish several quite separate issues about time. For a start, the authors overlook an ambiguity in the arrow metaphor, and hence confuse the issue of the directionality of time (that is, the question of whether the Universe is symmetric in time) with the question as to whether time flows. The standard use of the metaphor turns simply on the fact that arrows are effectively one-dimensional objects, with a clear orientation along this single dimension. But some arrows move

(unlike, for example, the ones in signposts), and this provides the trap that Coveney and Highfield fall into in saying that time "travels like an arrow" (page 24). Antiquity has given us a perfectly good metaphor for the flow of time, namely the stream or river. The arrow is best kept for simple directionality.

More seriously, the authors also fail to see an important distinction between the issues of directionality and determinism. This emerges early, when they misconstrue Einstein's remark that "the distinction between past, present and future is an illusion". They see this as arising from a "belief in a deterministic world" (page 30;

demolishes one of the main positive suggestions of the book. If chaos shows that the future is open, it also shows that the past is open. It does not yield an arrow of time.

Coveney and Highfield also suggest that the Brussels school's study of nonequilibrium systems provides a key to the resolution of the conflict between thermodynamics and mechanics. It does not. The real puzzle is why there is an arrow of time at all; that is, why the Universe is not simply a thermodynamic equilibrium at all times (except during the inevitable local fluctuations). The theory of nonequilibrium systems may tell us how such systems behave, given that there are some; but it does not explain how they come to be so common in the first place (and all oriented in the same temporal direction). This is "time's greatest mystery", and for all its merits, the theory of nonequilibrium systems does not touch it.

What would touch it would be a cosmological demonstration that the Universe was bound to be in a low-entropy state after the Big Bang. In their recent books Hawking and Penrose have each discussed proposals of this kind — approaches which at one point the present authors dismiss on the odd grounds that the cosmological theories concerned are "less firmly established than the phenomena to be explained" (page 34). Goodbye theoretical conjecture! The cosmological approach is at present inconclusive, but the popular works of some of its advocates — Paul Davies, for example, as well as Hawking and Penrose — provide the best current lay guides to the search for the arrow of time. Davies — that living testimony to the Universe's ability to create order at a prodigious rate — has also written a recent book

on chaos and nonequilibrium systems (*The Cosmic Blueprint*, Simon and Schuster, 1988). To my mind it has the edge on the present work.

Halpern's *Time Journeys* is subtitled "A Search for Cosmic Destiny and Meaning". We philosophers no longer work this territory, so I will not try to comment on whether the search is successful. The subtitle aside, however, the book is a modest but wide-ranging introduction to ideas about time in science and human culture. It would be useful introductory reading for an undergraduate course on this topic. There are many references to more advanced material. □

Huw Price is in the Department of Traditional and Modern Philosophy, University of Sydney, Australia 2006.



Father Time — still a mystery.

see also page 64). In fact the remark is simply an allusion to the familiar notion of a four-dimensional block universe, embodying no objective "present moment" or flow of time. Such a view need not insist that the state of the Universe at one time determines its state at other times.

In general, the issue of determinism is quite distinct from that of the directionality of time. A deterministic system may be asymmetric (for example, in virtue of its boundary conditions) and an indeterministic system may be symmetric. So even if it is true (by no means clear, in my view) that the theory of chaotic evolution "blows apart time-symmetric determinism" (page 206; see also pages 36–37), what it leaves might well be time-symmetric indeterminism. This

View from the north

Walter Gratzer

The Atheist and the Holy City: Encounters and Reflections. By George Klein. MIT Press: 1990. Pp. 200. £14.95, \$17.95.

THE riposte to a journalist's inquiry about life on Mars — that the Martians are already among us: they are called Hungarians — has been variously attributed to Arthur Koestler and Leo Szilard. Certainly the generation that grew up in Hungary at about the time of the First World War and the years following, succeeded in infiltrating much of European and American science. George Klein, who took root in Sweden, is of their number. He came to know Leo Szilard, physicist, biologist and political activist, as a friend and his vignettes of Szilard's life and character were for me the most rewarding parts of his collection of reviews and essays.

Szilard may have been something of a dilettante — one

who in the course of one revolving moon was statesman, fiddler, chemist and buffoon

— but it was he who first perceived the possibility of a nuclear chain reaction and at whose urging Einstein sent the celebrated letter to Roosevelt, that can be said to have launched the Manhattan project. Szilard was an inexhaustibly gregarious man and a tireless quidnunc: a law of physics had it that wherever two physicists met Szilard was created. He was moreover irrepressibly disputatious. The story (doubtless apocryphal) went that a jury on which he served in Chicago could reach no verdict, for Szilard was absolute for the accused man's guilt, the other jurors for his innocence. The judge invited the jury to retire and try again; they failed once more to agree, but this time Szilard found for innocence and the other eleven for guilt.

When Szilard became interested in biology his method of informing himself was to pay unannounced visits to biologists around the world and interrogate them minutely on their work, scribbling the while in his notebook. He would conclude the interview by holding out his pen with the peremptory demand, "sign please". And on then to the next laboratory on his circuit, like, as François Jacob put in his memoirs, "a sort of fat bumblebee, spreading pollen".

Klein too was subjected to these epiphanies, when Szilard would materialize at the door of his office and resume the conversation where he had left it the year before. Thereby too hangs a less familiar tale, for on one such occasion Klein followed his friend into the lavatory and noticed blood in the urinal. Szilard admitted when taxed that he had been passing blood for six months; he was too

busy, he said, to trouble himself about it. Klein took him to a urologist, who diagnosed a massive bladder tumour. Szilard mistrusted doctors; he rejected an operation, then accepted, changed his mind, offended the urologist and eventually opted, against advice, for radiation treatment at unprecedentedly high doses, and angles of irradiation that he himself determined. The cancer melted away and Szilard died in his sleep eight years later of a stroke.

Death recurs in these pages and indeed the book got started, it seems, in consequence of a commendation that Klein, as an expert on cancer, was asked to write for *In the Face of Death*, a reflection by a Swiss lawyer, Peter Noll, on his own impending demise. Noll had bladder cancer and his life could have been preserved by an operation, but he chose to let the disease take its course. Klein is deeply impressed by Noll's quietist approach to death — so different from Szilard's and from that of fighters, such as Peter Medawar, who gave so vibrant an account of why he preferred to stay alive. Klein suggests that these contrasted attitudes are represented in their extreme forms by the Japanese — who, he says, will not discuss mortality even between doctor and patient — and the Chinese, who look on the grim reaper's visage with the composure that Noll emulated and the Japanese envy. Klein asked an eminent and ageing Japanese oncologist to explain this yearning. After a long silence came the answer:

"Water birds fly over a lake. The lake surface, my soul. The birds' shadows the events."

Klein occupies a distinguished position in cancer research and appears to spend much time travelling: his experience of scientific affairs clearly goes well beyond the laboratory. The presence of a foreword by Lewis Thomas sets the seal indeed on his standing in the international scientific ascendancy. Thomas extols his academic lustre, and reveals — just to put the rest of us in our place presumably — that he is the author of 800 papers. Well, no matter. Klein has put his advantages to good account in these pieces. He is a sagacious observer of science and its practitioners; I enjoyed his essay on the social milieus that make for acceptance or rejection of new ideas. There are several good stories, such as the anecdote about Pasteur in old age: his young colleagues, observing the *patron's* habit of pausing from his work to mutter something to himself, wondered what the incantation could be that appeared to sustain him. One of his students drew close unobserved and this is what he heard: "*il faut travailler*". (Better by far than the story that clung to a famous Commodore of the Cunard line, who at moments of crisis would also soliloquise under his breath; the cabin-boy, deputed to eavesdrop heard: "port is left, starboard is right".)

Klein's dissertation on viruses and oncogenes is both enlightening and readable and his ruminations on the nature of evil and other riddles of the human condition are wholly sympathetic. Among the *belle-lettristes* of science he is not quite in the class of Jacob and Perutz, but his is a likable book with quite a number of buried truffles. □

Walter Gratzer is in the MRC Cell Biophysics Unit, Drury Lane, London WC2B 5RL, UK.



Sea turtle products — bags, a golf holdall and two stuffed hawksbills. Trade such as this, legal in many nations, accounts for thousands of deaths. *Fight For Survival* by R. L. DiSilvestro is published by Wiley, and focuses on the battle to preserve endangered species such as wolves, sharks, dolphins and others. Price £24.95, \$32.95. □

Remember then

Peter Harman

Physicists Look Back. Studies in the History of Physics. Edited by John Roche. Hilger: 1990. Pp. 393. £45.

THE study of the history of physics has become a flourishing area of scholarship. As a manifestation of its wider impact, in recent years some physicists have expressed interest in the educational value of the subject. The present volume reflects this burgeoning interest within the physics community, and collects papers presented to meetings of the History of Physics group of the Institute of Physics. As a result, this is a very miscellaneous collection of pieces, and while some bristle with facts and bibliographic data (David Hughes on meteors), others have the charm of the after-dinner speech (N. Kurti's reflections of an amateur historian). The editor has divided the papers as falling under three general themes.

The first part of the book is concerned with basic methods and includes papers on research libraries (Dennis F. Shaw), the techniques of oral history (D.H. DeVor-kin), and the resources and archives of contemporary physics (Spencer Weart and John Krige). Ronald Stansfield's common-sense discussion of the real problems raised by claims to 'repeat' past experiments is enlivened by personal reminiscences of the laboratory culture of the 1930s.

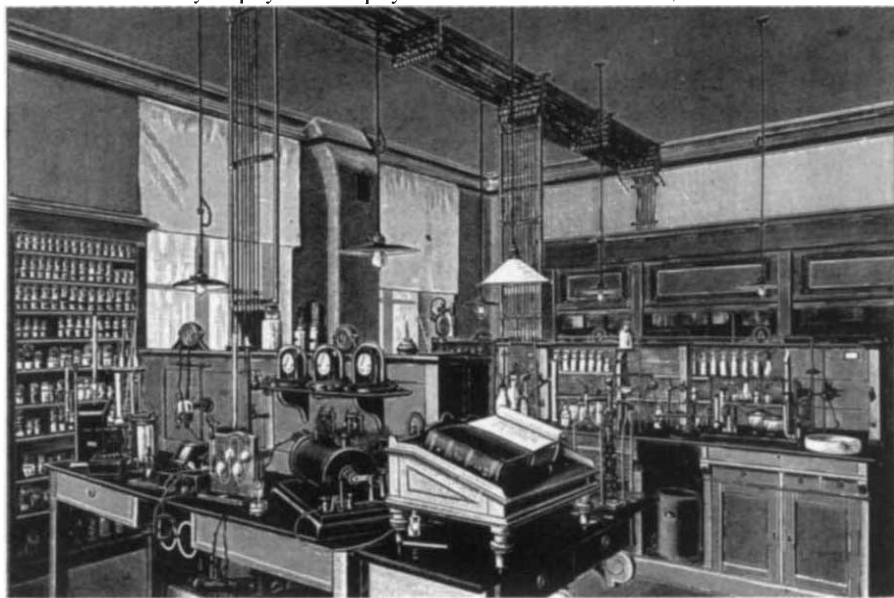
The second theme ranges over the various issues raised by claims for the educational value of the history of physics. A. P. French gives a cautious appraisal of the role of history at university level, while Brian Davies and Stuart Leadstone outline several approaches to the incorporation of historical materials within a school syllabus. Whereas these papers offer personal views on the ways in which history can enrich physics teaching, John Roche claims that history has a leading role in clarifying the conceptual foundations of physics. In exemplification of this theme he gives an account of the history of the concept of the vector potential from its origins in nineteenth-century electromagnetic theory.

The third theme of the book includes reminiscences of personal contributions to contemporary physics. Pieces by E. Lipson (crystal structure), D. J. Unwin (electron probe instruments), Sir Granville Benyon (atmospheric research) and B. D. Josephson (superconductivity) are included here. Of these papers, B. D. Josephson's account of the work which led to the award of his Nobel prize is likely to be of the greatest general interest. Josephson himself warns of the likely historical inaccuracy of his narrative, but such docu-

ments can form a valuable part of the historical record.

The problem of confronting physics with its history is discussed by several contributors. A.P. French quotes Pasteur and Sir Peter Medawar in illustration of the fundamental problem. As Medawar put it, "Science therefore in some sense comprehends its history within itself". What then can be the role for history? In his *Forces and Particles* (Macmillan, 1972), Sir Brian Pippard concluded that "If we are to eschew a truly historical presentation, surely it is best to avoid the pretence of historicity altogether". It may well be that history of physics and physics

no way pretends to be history as understood by historians, but draws upon their work to enrich the presentation. From the history side, Jed Buchwald's *From Maxwell to Microphysics* and *The Rise of the Wave Theory of Light* (Chicago University Press, 1985 and 1989 respectively), and Dan Siegel's *Innovation in Maxwell's Electromagnetic Theory* (forthcoming from Cambridge University Press), use the language of contemporary physics to describe and illuminate history, while remaining attentive to the integrity of the historical materials. This approach is not without its problems in characterizing historical nuances, but if the task of



Mary Evans

Laboratory culture — personal reminiscences enliven discussions of historical research.

are incommensurable. This would seem to be French's view, for he argues that "The presentation of history as historians understand it is close to being an impossibility in the physics class". Nevertheless, he believes that some historical materials can be used; these reduce to anecdotal vignettes, for we are engaged in 'the art of the possible'.

John Roche makes claim for a radically different view of history as essential to physics itself. Yet in seeking to give an account of the vector potential in its historical development as a prelude to fundamental conceptual clarification, his account epitomizes the problem as stated by French. Roche's analysis conflates concepts and styles of mathematics which were historically distinct and had different meanings in their original context.

The problems raised by Medawar, Pippard and French need to be confronted by physicists who wish to use historical materials and by historians who wish to address their work directly to physicists. On the physics side, Malcolm Longair's *Theoretical Concepts in Physics* (Cambridge University Press, 1984) incorporates historical material into a sophisticated undergraduate text, in a style that in

reappraising conceptual foundations is to be undertaken, this work offers a solid foundation.

In his preface, Roche divides historical research into three categories: sources, social history and technical history. He considers that the present volume falls into the latter category, but I consider this a misleading designation. *Physicists Look Back* does have its merits and interest, but it would be a pity if the physicists who are its likely readers came away with the impression that this is technical history as currently practised by historians of physics. Nor does the book give any indication of the scope of current work in the history of physics. To mention the study of sources alone, publication of the writings of the twentieth-century masters Bohr, Heisenberg, Pauli, and especially the Princeton Press edition of the *Papers of Albert Einstein* edited by Martin Klein and John Stachel, has been a major development during the 1980s. The editions of the works of the Bernoulli family and of Euler have recently taken on a new lease of life. Cambridge University Press are publishing Alan Shapiro's edition of Newton's optical papers, and D.T. Whiteside's recent facsimile volume of the *Principia*

manuscripts supplements his famous edition of the *Mathematical Papers*. Cambridge University Press has just published the first volume of my edition of *The Scientific Letters and Papers of James Clerk Maxwell*.

On a rather broader front, the literature abounds with interpretative studies, of which C. Jungnickel and R. McCormach's study of German physics, *Intellectual Mastery of Nature* (Chicago University Press, 1986), stands pre-eminent among recent publications. It would be a pity if physicists wanting to make use of historical work remained unaware of all this effort. □

P. M. Harman is a visiting scholar in the Department of History of Science, Harvard University, Cambridge, Massachusetts 02138, USA.

Holy grail or snark?

J. Ellis

The Higgs Hunter's Guide: Frontiers in Physics. By J. F. Gunion, H. E. Haber, G. Kane & S. Dawson. *Addison-Wesley*: 1990. Pp. 425. £41.35, \$49.50.

WHAT is the most pressing problem in particle physics today? All confirmed experimental results agree dramatically with the Standard Model, giving us no clue how to progress beyond it. But the Standard Model is very cumbersome and inelegant, containing three distinct sets of gauge interactions, nine *a priori* arbitrary quark and lepton masses, four assorted mixing angles and phases to describe the charged-current weak interactions, and two other parameters thrown in by hand to break spontaneously the electroweak symmetry group. The latter 15 parameters all describe the couplings and mass of the mythical Higgs boson, supposed purveyor of masses to all the other elementary particles, but as yet unseen by experiment and the subject of furious theoretical debate. These seem reasons enough to cite the quest for the Higgs boson as the most pressing problem in particle physics, though some might cite the unification of all the interactions and understanding the proliferation of different particle types (the "flavour problem"). However, I consider the origin of mass and the mechanism of symmetry breaking prerequisites for addressing these important problems, particularly because the reasons for thinking that the Higgs boson should be accessible to present or planned accelerators are better than those for expecting detectable manifestations of unification or flavour physics.

With the recent start-up of the Large

Electron-Positron Collider (LEP) and advanced plans for the Large Hadron Collider (LHC) accelerator in Europe and the Superconducting Super Collider (SSC) in the United States, this book reviewing strategies for detecting the Higgs boson is therefore particularly timely. Moreover, the authors are among the best qualified to write on this important topic. Indeed, they provide the reader with a large amount of useful and accurate information about the expected properties of the Higgs boson and how to search for it in e^+e^- and hadron-hadron collisions. Nor is their attention focused only on the minimal Standard Model, because much of the book discusses non-minimal Higgs sectors, such as appear in supersymmetric and technicolour models. There are also discussions of axions, majorons and composite models. Also very useful are lengthy appendices containing detailed Feynman rules and formulae for Higgs decay rates. There is no doubt in my mind that this book should find a place in the office of any particle physicist interested in the problem of spontaneous symmetry breaking, which should include everyone working in high-energy physics.

I have only a few minor quibbles with the content of the book. It would have been good to have had a more extensive and elementary introduction to the motivation for spontaneous symmetry breaking in gauge theories and a general discussion of the Higgs mechanism. Some comparison and contrast with the theory of superconductivity could also have been appropriate, and might have made this excellent book more accessible to students and nonspecialists. Likewise, some discussion of the "naturalness" or "hierarchy" problem as a motivation for supersymmetric and technicolour models would also have been welcome.

As is perhaps inevitable in a topical book such as this, it has to some extent already been overtaken by events. Much of the discussion of a possible light Higgs boson has been rendered moot by the success of LEP at CERN, where just the Aleph experiment alone has already excluded any Standard Model Higgs boson weighing up to 41 GeV, and the combined data of all the four LEP experiments probably exclude the entire mass range up to 49 GeV. LEP has also given many limits on nonminimal Higgs bosons. By comparison, in their figure 1.1, the authors predicted that LEP and the Stanford Linear Collider should be able to reach a mass of 30 GeV only in 1993. Unfortunately, the SLC has been unable to muster sufficient luminosity to have any significant impact on the search for the Higgs boson.

These minor defects could easily be corrected in a second edition and do not detract from the great worth of this book,

which will be an invaluable handbook until experiments find the Higgs boson or whatever replaces it, and probably long afterwards as well. This book will surely serve its avowed purpose of guiding the Higgs hunter towards her or his prey, and also interpreting whatever game she or he succeeds in bagging. □

John Ellis is in the Theory Division, CERN, CH 1211 Geneva 23, Switzerland.

Lucky dip

Stuart Sutherland

Images and Understanding. Edited by Horace Barlow, Colin Blakemore and Miranda Weston-Smith. *Cambridge University Press*: 1990. Pp. 401. Hbk £40, \$69.95, pbk £15, \$24.95.

IN 1986 the Rank Prize Funds financed a conference on the use of images. Because I forgot to go, I can comment freely and without bias on the proceedings, now published somewhat tardily by Cambridge University Press under the title *Images and Understanding*.

The conference was nothing if not interdisciplinary for there were papers by psychologists, neurophysiologists, philosophers, computer scientists, and for good measure a writer, a theatre producer and a choreographer. Despite all the evidence refuting the belief, today even scientists seem to think that if enough people are gathered together, something good is bound to emerge. In this case what emerges is less a treasure trove than a lucky dip, which contains few pearls but has many trinkets whose glitter may amuse for a while.

Perhaps the most original contribution is that of John Willats who combines a practical knowledge of the visual arts with a formal training in engineering and a somewhat less formal one in psychology. He notes that paintings can attempt to represent surfaces either as seen from the observer's viewpoint or as they really are in 3-D space. Perspective is the prime example of the former category. An example of the latter is orthographic projection, in which selected surfaces are shown in their true physical shape, giving object-centred descriptions of at least part of the scene. Thus in P. Bonnard's picture, *Nude in a Bath Tub*, the top and one end of the bath are depicted in their true shapes, rather than in the shapes they would project to the retina. Willats also categorizes pictures and drawings by their denotational system, that is the rules or conventions for interpreting the marks in the picture. In a denotational system that is viewer-centred the brightness and colour of the marks reproduce an image on the retina that approximates as closely

as possible to that which the scene would produce when viewed naturally. In other denotational systems, such as cartoons, a line may represent an edge: many other variations in brightness, for example those produced by shadows or by changes in albedo may be omitted. In some systems, hatching may arbitrarily represent a shadow. As Willats points out, such denotational systems have preprocessed the scene by making explicit the features that produce changes in brightness such as an edge or a shadow. These are novel and interesting ways of categorizing pictures, and they illuminate the motivation underlying different styles of painting.

Three other papers presented ideas and findings that were new — at least at the time of the conference. They were on colour (John Mollon), motion-detecting cells (Tony Movshon), and cells in the monkey's inferotemporal lobe that respond maximally to a person moving in a particular direction — backwards, forwards, left or right (David Perrett). Interestingly, some cells respond to directions of motion relative to the moving person's own position: one cell responded maximally to a person moving forwards to the right, but responded much less if he was facing left and walked backwards. If these results hold up, we are coming perilously close to grandmother cells. The interesting question is not, however, whether such cells exist, but how, if they do exist, they are wired up. The silence elicited by this problem is as profound as the problem itself.

To turn to the trinkets, the one that glitters most is a paper by Jon Darius on the fallibility of scientific observations, particularly in astronomy. It is well known that Martian canals were only disposed of by photographs from space craft, but he gives several other instances in which astronomers have collectively seen patterns that do not exist. One nice story concerns the map of Mercury that one astronomer compiled when it was thought that Mercury took the same period to rotate as to complete one orbit, thus always presenting the same face to the observer. Although it is now known that these periods are quite different, three other astronomers, each looking at a different face, drew maps that were almost identical to the original one.

Of the other papers, two may be singled out for their amusement value. John Krebs recounts the behaviour of the honeyguide, an African bird that is adept at locating bees' nests but rather poor at obtaining access to them in order to eat the honey. It has solved the problem by leading human hunters to the nests it has located: when the hunters open them up, it is able to gorge itself. Mike Land introduces us to the scallop's unusual method of forming a retinal image — by a concave mirror behind the retina — and

the even more bizarre method used by the shrimp — a series of flat mirrors in front of the retina all of which send light from the same point in space to the same point on the retina: he cites several similar curiosities. Entertaining though such discoveries may be, they are hardly likely to change the way scientists think.

In fact, there are few theories, whether grand or small, to be found in this collection. Too many of the authors tell us, often in some detail, what we already know. Colin Blakemore puzzles over the question of why visual and other sensory maps exist, but misses the obvious answer that normally it is necessary to analyse the retinal image locally. The parts of an object are contiguous and must be processed together. In a different world, it might have been necessary to put together the pattern of light falling in the top right-hand corner of the retina with that in the bottom left, which would have needed a form of mapping quite unlike existing cortical maps.

The book has two curious features. The first is what it leaves out. There is nothing

on the difficulty of interpreting medical X rays or the electrocardiogram nor on the merits of attempting to do so by computer. Nor is there anything on the bizarre disruption of images that can occur in brain-damaged patients. Dreams and hallucinations receive no mention, and apart from a philosophical contribution there is little on the nature of mental images. Second, the individual papers have remarkably little relevance to one another. It is just possible that some of the participants will be inspired by the contributions of workers in other disciplines to change their view of their own patch, but it seems unlikely.

Most of the papers are clear and some are elegant: few are deep. They will appeal less to the scientist than to the general reader, whose delight in trinkets is unaffected by their antiquity. I half wish I had remembered to attend the conference. □

Stuart Sutherland is at the University of Sussex, Laboratory of Experimental Psychology, Brighton, BN1 9QG, UK.

Pretty pictures

P. T. Saunders

Envisioning Information By Edward R. Tufte. *Graphics Press: 1990. Pp. 126. \$48.*

ONE picture, so we are told, is worth a thousand words. Certainly most scientists seem to think so, judging by the large number of figures and tables we put into our papers. With pictures even more than with words, however, it is not always easy to convey information in a way that is both attractive and clear, especially because in both cases the medium is less complex and dynamic than the message it is being asked to carry.

In *Envisioning Information*, Edward Tufte discusses how pictures can be made more effective. But the book is not a text on graphic design. Its aim is to illustrate general principles that help to identify and explain design excellence, not to teach specific techniques. Nor is it aimed specifically at the scientist. There are, to be sure, examples from science, but more from other fields. But they have been chosen with a view to holding the reader's interest, and almost anyone is bound to come across something that he did not know before and is likely to remember.

Tufte provides many beautiful examples of "cognitive art" — and a few horrors as well, to show what can happen when the principles are ignored. The illustrations include such gems as diagrams from an 1847 colour edition of Euclid's *Elements* in which the figures resemble paintings by Mondrian. And even though the focus

of the book is on graphics, the text is especially well written.

All the same, *Envisioning Information* is likely to disappoint any scientist who buys it chiefly in the hope of learning how to write better papers. After reading it, I went through a recent issue of *Nature* looking for figures that I thought would have been better if the authors had followed some of Tufte's suggestions, and I did not find any. That may reveal a lack of imagination on my part, but I think it also indicates that most figures in scientific journals are pretty straightforward. Most of the graphs looked clear enough to me, and I doubt there is very much that even Tufte could do with a photograph of a gel.

A particular shortcoming from the point of view of the scientist is that much of the book is concerned with the effective use of colour. This is an important topic, and it makes the book itself more pleasing to look at, but it is of little practical value for authors who are almost always restricted to black and white on grounds of cost.

Envisioning Information is not as relevant for the scientist as Tufte's earlier *The Visual Display of Quantitative Information*. You would not want to keep it on the laboratory shelf next to handbooks of constants, tables of integrals and style manuals. On the other hand, I enjoyed reading it, it did teach me something about the use of graphics — and if it had been a more practical text I would not have learned anything at all because I would never have picked it up in the first place. □

P. T. Saunders is in the Department of Mathematics, King's College, The Strand, London, UK.

Conflict and concern

David Victor

Energy Policy in the Greenhouse: From Warming Fate to Warming Limit. Volume 1. By F. Krause, W. Bach and J. Kooney. *Earthscan, London: 1990. Pp. 224. £29.95.*

Global Warming: The Greenpeace Report. Edited by J. Leggett. *Oxford University Press: 1990. Pp. 554. Hbk £19.95, pbk £5.95.*

RECENT additions to the burgeoning supply of books on global warming have generally offered few novel insights. General readers will find these two new studies of the science and policy issues adequate; experts in the field should recognize virtually all of the arguments. Readers sympathetic to the underlying argument of both studies — that policies should rapidly be set on the risk that climate change will be extreme — will find support for their position in these studies. Critics may be annoyed by the abundance of unsupported statements and speculation.

Despite its title, *Energy Policy in the Greenhouse* devotes little attention to energy policy. Instead, it is an overview of various issues related to global warming, emissions of greenhouse gases and assorted policy options.

Most other studies follow the tradition of energy modelling and start with assumptions about energy use, calculate greenhouse-gas emissions and then predict future warming under various possibilities. Refreshingly, *Energy Policy* takes the opposite approach; the authors start with a conservative "warming limit" from which they compute acceptable patterns of greenhouse-gas emissions.

The other main feature of this study is the integration of geophysical and greenhouse-gas emissions models. Notably, the authors present a persuasive case for examining temperature increases over time — known as the transient climatic response — which they argue should be limited to a tenth of a degree centigrade per decade (a reasonable number based on the rate at which trees can adjust to climate change). Only a few other studies have explicitly linked the transient response of the Earth's climate to energy; as such, this research is innovative.

Unfortunately, the report is poorly integrated. Alternative scenarios that are laboriously developed and analysed early in the volume have little relevance to the policy recommendations which emerge later. In all, I have been able to identify at least 28 distinct possibilities which appear at various points in the study. Especially unfortunate is the lack of continuity between the geophysical modelling and the emissions modelling. Furthermore, innovative consideration of the transient response dissipates after the first third of the book. In the final chapters, the range of scenarios the authors deem acceptable

is dominated by static conceptions of warming and does not reflect the limits and opportunities of the transient response.

Throughout the book, the authors posit incisive questions yet answer these queries without analytical rigour. The policy recommendations span a wide range of industrial and agricultural practices, but the report sacrifices depth in its attempt to be comprehensive. (But readers should look forward to volume 2, currently in preparation, which focuses more sharply on energy policy.)

The *Greenpeace Report* is a collection of nineteen essays organized around the science, impacts and policy responses related to global warming. Readers will recognize these as the same organizing principles of the three working groups of the United Nations-sponsored Intergovernmental Panel on Climate Change (IPCC) which has been preparing a consensus report on global warming for the past two years. The Greenpeace view of global warming is framed in a critique of the IPCC report.

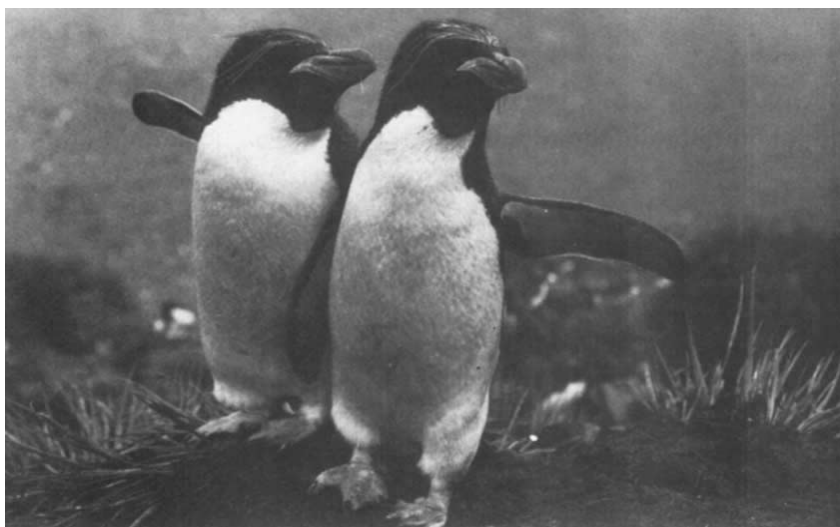
Greenpeace argues that IPCC has (1) failed to make the risks of climate change apparent, (2) systematically understated the effects of climate change and the risks of extreme change, and (3) failed to recommend stringent measures to slow and reverse global warming. This third

attack on IPCC's policy recommendations forms the bulk of the study.

The quality of the *Greenpeace Report* is uneven. Several of the essays are notable, including those by Birgit Bodlund *et al.* on Sweden's electricity sector and by Norman Myers on tropical forests (both of which are versions of studies reported elsewhere). Amory Lovins provides his familiar but compelling case for least-cost planning, arguing that it is cheaper to save energy by increasing efficiency than it is to purchase new energy sources. According to the argument, controlling global warming can be profitable.

Some interesting hypotheses emerge from least-cost planning and the data presented by the Greenpeace authors. For example, Bill Keepin repeats his claim that carbon emissions can be reduced at 0.012 tons per dollar invested in energy efficiency, whereas building nuclear plants to offset carbon emissions only displaces 0.002 tons of carbon per dollar. But Myers' numbers suggest that reforestation offsets carbon at an even more efficient 0.75 tons per dollar. Accordingly, every dollar invested in energy efficiency could therefore displace six times the carbon if it were invested in reforestation.

Based on Keepin's data, the concluding chapter of the *Greenpeace Report* argues that: "Expansion of the nuclear-power industry . . . would involve diverting funds from cost-effective energy-solutions to the greenhouse crisis . . . [and] should have no role to play in the international policy response." Does Greenpeace believe that (considering Myers' data) investments in energy efficiency should have no role in the policy response because they divert funds from reforestation? Presumably the political answer is no, although their economic data may



Double greetings — A pair of macaroni penguins on their tussock mound nest site in the Welcome Islands of South Georgia. *Wild Ice: Antarctic Journeys* by R. Nareen, C. Monteath, T. de Roy and M. Jones is a desk-top voyage to the Antarctic, illustrated with stunning photos of the wilderness and the birds and animals there. Published by Smithsonian Institution Press, price is \$29.95, £19.95. □

favour such an extreme conclusion. What is missing from the synthesis is a close study of which policy actions are more important than others, given that the economics of greenhouse policy conflict with other social priorities.

Clearly there are empirical gaps in the climate policy literature, especially regarding financial cost. Thus, a first step in an international programme to slow global warming should probably include, for example, a diverse programme of reforestation to test Myers' data rigorously under real conditions. This research could be pursued in the near term without a full-blown climate agreement.

Beyond the wide range of scientific questions that remain unanswered, equally challenging is the need to develop effective policy measures which consider national styles of governance. Good greenhouse policy for one nation may not be appropriate for another, in light of disparate cultural, socioeconomic and environmental conditions; useful policy analyses of global warming must tackle these pragmatic concerns. For example, it

is not sufficient for the Greenpeace authors to note that methane (a strong greenhouse gas) could be reduced by feeding cows better food. Instead, policy recommendations must go further and provide workable approaches to the salient, practical questions such as: how do we convince a North American cattle rancher and an Indian herdsman to feed their cows better food?

The political will which has coalesced around global warming is admirable, but urgent calls for immediate policy action must recognize that (1) relevant science and policy fields remain underdeveloped, and (2) concern for the greenhouse should not eclipse other pressing problems of our time. As both studies argue persuasively, these factors should not paralyse policy response. But severe climate change is not the only risk; misdirected policy also poses a serious threat. □

David Victor is at the Department of Political Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Portrait of a dinosaur

Seth Shulman

Seabrook Station: Citizen Politics and Nuclear Power. By Henry F. Bedford. *University of Massachusetts Press: 1990. Pp.224. \$22.95.*

IN the decade since the partial meltdown of the Three Mile Island Unit 2 reactor in Pennsylvania in 1979, the construction of new nuclear power plants in the United States has ground to a virtual halt. No new reactors have been ordered since the accident and all of the 47 units ordered since 1974 have ultimately been cancelled.

During this ignominious period for the US nuclear industry, the Seabrook nuclear plant in New Hampshire has stood as a powerful symbol of the faltering status of the industry — and of the wide-spread domestic opposition to nuclear power. Seabrook has now been licensed to operate, but angry critics and stalwart nuclear proponents alike rightfully see the process as a tragedy.

In 1967, when officials at a tiny New Hampshire public utility got the idea to build Seabrook, they planned two 1,100-megawatt units at a projected cost of \$425 million each. Now, in 1990, the utility has only marginally survived the two-decade ordeal. It has gone bankrupt. It has managed to build only half of the plant envisioned initially. The plant's construction has divided communities across New England, the process has cost the utility — and ultimately ratepayers — an

astronomical \$6,400 million, an amount that will be nearly impossible to recoup no matter how high the cost of oil.

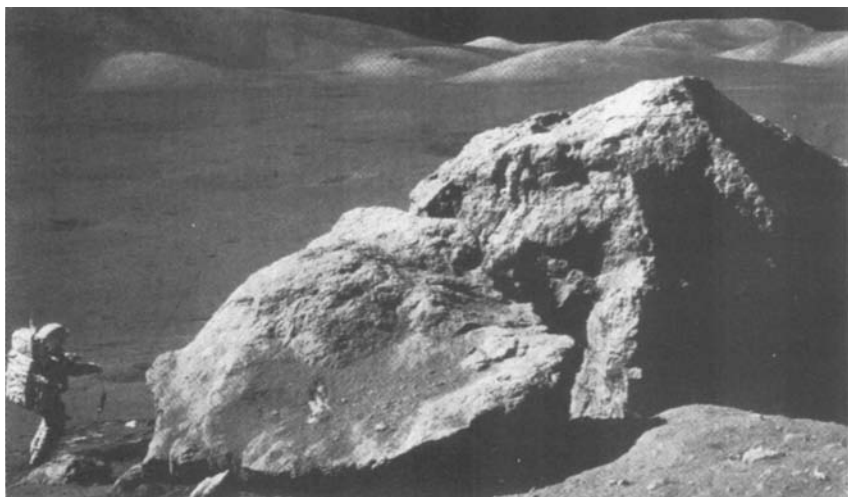
Bedford's book plumbs the depths of this remarkable story. This is not a referendum on the merits or hazards of nuclear power, but rather a lively, thoughtful and dispassionate look at a tangled and dysfunctional regulatory process. A book about bureaucratic politics may sound dull, but this one could never be so accused. Not with the high drama Bedford is able to chronicle; not with the cantankerous cast of characters involved every step of the way; and not with the nuclear hopes of so many tied to the success of the Seabrook project.

At face value, Seabrook's story is

fascinating itself, but it is Bedford's analysis of the interminable proceedings that makes the book special. He offers a remarkably sympathetic portrait of each side of the struggle over Seabrook, but also shows how each repeatedly abused the system. Much of the debate in the later years of the licensing struggle at Seabrook revolved around the issue of evacuation in the case of an emergency. As Bedford shows, for many months the entire fate of this completed reactor rested to a large degree on whether, in the event of an accident, a school bus full of local residents would overstep a particular town bridge, thereby jeopardizing the planned evacuation process.

By most accounts, it is now clear that Seabrook's siting was a blunder. Poised on the edge of a crowded summer beach resort, the possibility of safely evacuating the summer population in the event of an accident is improbable at best. And yet the debate over siting was not allowed to take place in a meaningful way until the reactor was almost finished — hardly an ideal arrangement for either side. The residents, a great number of whom simply did not want Seabrook built in any form, stymied licensure by raising issues like the health of fish and the safety of bridges. These were ploys to stall — and hopefully stop — the plant altogether. Meanwhile, though, with so much money and vested interests resting on the plant, Bedford shows how the regulatory process was turned into a sham. The state and federal officials involved effectively rewrote almost every rule governing public safety to ensure the plant's operation. In essence, the process could be delayed, but the powerful players involved were not going to let the fate of an enormously expensive project be foiled by a bureaucratic technicality.

There is an important lesson here. As



Man on the Moon — Harrison Schmitt, geologist and pilot of the Apollo 17 Lunar Module, examines an enormous broken rock on the Moon in 1972. *The Cambridge Encyclopedia of Space* edited by M. Rycroft is "the most complete and up-to-date account of the conquest of space" published in one volume. Published by Cambridge University Press at £40, \$65. □

we learn in some detail, the US licensing process allowed groups to quibble for more than twenty years about the solidity of bridges, and the minuscule rise in ocean temperature caused by the plant; at no time in the process were people able to debate in a meaningful way the overarching issue of nuclear power, or the political and philosophical underpinnings of the participants that led them to such divergent views. Woefully, it seems, our society has yet to learn the process or vocabulary for such debates. Hence, technological questions of major importance to us all seem to get decided, if at all, in a shamefully *ad hoc*, *de facto* way.

If there is a flaw to Bedford's careful and clear case study of the Seabrook plant it is that he does not provide the reader with quite enough context fully to appreciate the circumstances. Most notable, of course, is the parallel story of Seabrook's sister reactor, the Shoreham plant in Long Island, New York. The unopened Shoreham plant has now been sold to New York state for one dollar, with the state assuming closing costs and vowing that the plant will never generate electricity. The divergent ends reached in these two cases make for a fascinating comparison: why did one constructed plant receive its licence whereas the other was effectively blocked by opponents? Similarly, Bedford could have offered more general background of the status of nuclear power in the United States. On economics issues, for example, he admirably explains the complexity of the financing as Seabrook's small public utility went ever more deeply into debt. But Bedford might have offered a bit of context here as well: all of the several lingering nuclear plants under construction during this period faced skyrocketing costs. Such a comparative perspective might have reminded us of the broader world in which Seabrook operated.

But this was not Bedford's aim. Rather he sought to do a close portrait of one dinosaur, and as such his work succeeds splendidly. In the end we learn much about how not to conduct public debate about technology. But we also learn how so many poor decisions could have been piled on top of each other to allow such an ill-advised project to proceed. The reasons are complex, but they are also caricatured beautifully by an administrator early on in Bedford's book advancing his so-called dinosaur theory to explain the Seabrook phenomenon. As Bedford writes, the administrator asks, "If you have a dinosaur in your backyard, you have to keep feeding it, because what can you do with a dead dinosaur?" We have only to watch the nearby Shoreham plant, I suppose, for answers. □

Seth Shulman is at 28 Bates Road, Watertown, Massachusetts 02172, USA.

Indigestible science?

Ken Ducatel

The New Technology. By Dimitris N. Chorafas. *Sigma: 1990 (marketed by John Wiley and Sons Ltd). Pp. 410. £14.95.*

THERE have been many recent books which have brought advances in science to the attention of a wider audience. Current media interest in issues such as the new physics, chaos theory, cosmology, genetics and artificial intelligence undoubtedly originates in popular science books written by specialists in these fields.

What we still lack is an accessible, but properly documented, overview of the implications of advances in science and technology for medium to long-term industrial strategy. With the widespread recognition of the salience of new technologies in competitive performance such a book has a guaranteed market amongst industrialists and policy makers. Yet it does not exist. It seems that the main reason why this book has still to be written is that it would be incredibly hard to do well. Not only would it have to summarize the technologies quickly, simply and accurately, it is also notoriously difficult to forecast technological progress accurately.

Good authorial style would be crucial. It takes a certain skill to make science digestible to a general audience whilst not being howled down by the experts in the field. It would also be necessary for instance to observe scientific rules about the presentation of information, such as adequate attribution and accuracy in the use of data, as well as extensive use of visual devices and conceptual models to convey the substance of the sometimes elegant, sometimes complex subject matter.

We may also ask, how confident an author can be in identifying the precise areas of technical advance which will be economically important in the next ten to twenty years. It is hard enough to forecast with precision how fast single new technologies are gaining acceptance in the commercial world. When it comes to describing the interaction between economic forces and an array of new technologies, accurate forecasts are out of the question. The only solution to the forecasting problem is a powerful explanatory model of the process by which new systems of technology are introduced into society. In the past there have been periods of great economic change and crisis as an existing system of technology burns out and is shaken out of the economy and the new technological system is installed. We have been undergoing just

such a period of shift in the "techno-economic paradigm" since the early 1970s. The rate of adoption of new technologies and therefore their significance to industrialists and policy makers will depend upon their relation to the point in the cycle of these paradigms which we have reached. By using a powerful concept such as the techno-economic paradigm, order can be imposed on the book's analysis of the economic and social consequences of competing science and technology developments.

A full plan for the book on future technologies for concerned managers cannot be worked out here, but its fundamental features are given above. It would have to be written in an accessible but reflective style, so that justice is done to scientific knowledge. It would also need a strong conceptual core, to give the reader an orientation to the overall streams and currents of technological change. Unfortunately, although apparently recognizing that the niche exists, Dimitris Chorafas has produced a book which fails to meet these criteria. I will restrict my discussion to three basic problems: style, attribution and structure.

First, the writing style of the author is idiosyncratic to say the least. In what is apparently an attempt to write in a down-to-earth style, he uses short choppy sentences, sometimes dispensing altogether with the verb in the sentence. There are instances where the author's use of English is so strange as to make the text incoherent and in some cases laughable. A howler such as "winds of change can be seen" (page 15) is merely an example. The general effect is of a rather rushed translation by someone who is not a native English speaker. Second, the book lacks any systematic reference system; despite the author's habit of copious name dropping, few references are listed. The lack of attribution in popular nonfiction is justified on the grounds that full referencing obstructs the flow of argument. In this book it is symptomatic of a general neglect of appropriate standards in the reporting of scientific knowledge. Readers not familiar with an area who want to follow up points which are being made in the text will find this impossible, as the book lacks even a guide to further reading. Graphs and tables suffer a similar fate with, in most cases, no attributed source and in some cases no scale on the axes.

Third, the book lacks a central conceptual approach, beyond the theme that 'science is discovery is progress'. As a result the discussion of the different technological developments is without focus. The book styles itself as a survival guide for the 1990s, which would imply that it should be used as a resource to gain access to some of the leading-edge technological developments. In the absence both of a systematizing conceptual model of how

technologies move into the economic sphere and without an adequate referencing system it is hard to see how the book can be termed a guide in any sense of the word. It cannot be used to gain access to a wider world of knowledge about specific technologies, as in a hitch-hikers guide to new technology. Nor can it be used as a knowledge development tool in its own right, in the sense of a do-it-yourself action plan to new technology.

Somewhere in this volume is an interesting book trying to get out. The hyped-up, chopped-down delivery style appropriate to managerial seminars does not work in book format. Although the author may have a grasp of a wide range of recent scientific developments, his ideas about what these mean for the future are not effectively conveyed here. Perhaps the management seminars are better value. A treatment of these issues aimed at policy makers would be very appropriate during the current period of technological restructuring. It is a shame, but this is not the book. □

Ken Ducatel is in the Department of Science and Technology Policy, University of Manchester, Oxford Road, Manchester M13 9PL, UK.

Hidden message

John Galloway

Opportunities in Biology. National Academy Press: 1990. Pp. 448. \$51.

THERE is a tale — I hope it is not apocryphal — that Sir Lawrence Bragg, asked by the BBC to take part in a radio programme, refused — on the grounds that he did not think the country could afford to have a significant part of its scientific manpower out of action for an afternoon.

No such anxiety appears to have assailed the compilers of *Opportunities in Biology*. The book was put together by a committee of 20 of the most successful biologists in the United States, including luminaries such as Maxwell Cowan, Leroy Hood and Eric Kandel. By way of explanation, the preface says "no single individual can hope to grasp all the new activities and opportunities (of biology)". Nor could even 20 apparently, who were aided and abetted by a National Research Council staff of seven, eleven expert panels with at least six members each, a Board on Biology, a Commission on life sciences and 94 other contributors and reviewers.

What was it Horace said about the mountains labouring? Here we have not merely any old mountains but the Olympian peaks themselves, a couple of hundred of America's finest. Horace's mountains brought forth a mouse. What has the National Research Council been

delivered of? The answer is not little, but is, I am sorry to say, of little value.

Apart from its subject, you judge a book by the quality of its language. That this is a book about biology, about ideas and methods that we can confidently expect to be unfamiliar to many of those at whom it is aimed — administrators and policy makers, for example — makes the proper use of language essential. And in its language, *Opportunities in Biology* leaves pretty much everything to be desired. It is remarkably sloppily written — it is not precise, economical, vivid, concrete and straightforward; it does not give sufficient apt detail to support the points it wishes to make; numbers are prominent only by their scarcity. The writers have included some graphic illustrations but appear to have chosen them at random without a decided or consistent purpose.

Let me give one or two examples of language used sloppily or wrongly — a few among many:

The word environment means surroundings. It does not need the adjective surrounding to precede it.

Epstein Barr virus (EBV) does not "carry" cancer. The virus is probably one of a chain of causes leading to Burkitt's lymphoma or nasopharyngeal cancer. (It is likely that EBV infects 90 per cent or more of the world's population but the cancers in which it is implicated occur in very few areas of the world.)

On page 43, we are told that the largest single DNA that has been sequenced is the genome of EBV with 172,000 base pairs; and yet on page 46 that the genomes of organisms range in size from 750,000 to 3 billion base pairs. Whether or not a virus is an organism depends on your view point. But to read those two statements three pages apart without a word of explanation or qualification is simply unsettling. Incidentally, the longest single piece of DNA sequenced is now that of cytomegalovirus with 230,000 base pairs. This sequence, completed last year at the Laboratory of Molecular Biology in Cambridge, is only 30 per cent longer than that of EBV. This is worth mentioning because it suggests things are progressing relatively slowly in terms of sequencing complete DNAs. The optimistic note struck in *Opportunities in Biology* in writing about sequencing the human genome might have been tempered by this fact.

That the book is slightly out of date in respect of the cytomegalovirus genome is probably not very important. But it is out of date in more significant ways. Arguing for more research funding on the grounds that the United States is slipping against other countries in the quality of its biological research, figures of the number of US papers in the top 10 per cent worldwide are used. But the best that can be

done apparently is to use data comparing 1973 with 1980. Data one to two decades old in a book like this are not good enough.

Picking up on Sir Lawrence's view point, I could not help irreverently wondering if one reason for the worsening performance of the United States is the extent to which its scientists are involved in exercises like this one.

Actually, I was almost more concerned with the quantity of language in the book than its quality. A large proportion of it seemed superfluous. And indeed genomes seem to offer a good way of illustrating the fault. It is usual to offer language as an analogy for the information contained in DNA. In viruses all the DNA is involved in the genetic message which is precise and compressed. On the other hand, in the human genome, 97 per cent of the DNA seems to have no purpose. The genetic messages are contained in about 3 per cent of the DNA. I could not help feeling that human DNA provides a useful analogy for language as it is used here. The messages in *Opportunities in Biology* are contained in a small fraction of the total words (the rest just has to be ploughed through).

The book also seems curiously muddled, both in small and large ways. The executive summary tells us of some of what biology has achieved by way of understanding. It then tells us "it is ironic that a time so filled with great opportunity should also be a time when a major fraction of the diversity of life on Earth is in danger of extinction". It is not ironic. The two facts are tightly tied to each other. The main problem the Earth faces is overpopulation by people and over-use of resources by a proportion of them, the end product of the "successful" application of scientific knowledge in the form of agriculture and medicine, for instance. I could not find a section about overpopulation, nor does it have an entry in the index. Where it might have been is the word ostrich — that does seem ironic.

What is strange about the flaws in the book is that 20 years ago the same exercise was undertaken in a volume called *Biology and the Future of Man* edited by Philip Handler, president of the National Academy of Sciences. It was very much better written and better at explaining some of the fundamental science. And it did conclude that population was the major problem for the immediate future. The earlier book was published just before two momentous pieces of biology — the discovery of a way of making monoclonal antibodies and the development of the technology for sequencing DNA — which between them changed the face of biology almost beyond recognition. I would have thought that fact alone would have caused the writers of the present book to remind its readers that predicting what will happen is a somewhat uncertain exercise in science as elsewhere.

I think if the National Research Council is thinking of doing this again it ought to attempt something more limited. The book is arguably an exercise in the public understanding of science, which has three different sides to it — its ideas and methods, its scope, and its relationship to society, government, industry, medicine and agriculture. A book beginning with a range of problems that face mankind and addressed by considering the scope of modern biology and how research is turned into action — and what sort — by industry or government would be extremely useful, or at least could be. To mix it up with attempts at Mickey Mouse explana-

tions of protein crystallography and the rest of molecular biology is simply not very helpful. Furthermore, books constructed on the basis of panel discussions need the iron hand of a very determined editor to lick them into shape.

But why not commission someone to write the book for them? A decent science writer could do this without the need for a huge formal structure of committees and panels — and I am sure Sir Lawrence would have approved. □

John Galloway is at the Cancer Research Campaign, 2 Carlton House Terrace, London SW1Y 5AR, UK.

Moral dilemma

Philip Kitcher

Created From Animals. By James Rachels. Oxford University Press: 1990. Pp.245. £17.50, \$19.95.

FREDERICK Banting and Charles Best spent most of the summer of 1921 in a small basement room at the University of Toronto, experimenting on dogs. Out of their work came the discovery of insulin, relief for millions of diabetics whose lives have been enormously prolonged through daily injections. Some of those who lived gave birth to children who would never have been conceived but for this discovery. I was one such child, one of the many who owe their lives to Banting and Best.

The work that led to the discovery of insulin involved considerable human sacrifice and suffering. Banting and Best were ultimately rewarded, primarily with the satisfaction of having made a remarkable medical advance, secondarily with high honours in the scientific world. But other actors in the drama were not so fortunate. Dogs do not win Nobel prizes. They are "sacrificed" for the advance of human science.

Can such sacrifices be justified? Are human beings morally entitled to use nonhuman animals to advance human welfare? Reflecting on the discovery of insulin, many people — especially, perhaps, those whose lives were made possible by it — will regard the question as frivolous. It is sad, of course, that the dogs in the Toronto basement had to suffer, sad but necessary. When we look at the great good that has ensued, the sufferings, the deaths, are justified. Would we say the same if Banting and Best had operated on people, if they had coerced adults against their will, or used orphaned children, or performed their operations on severely deformed neonates? In those circumstances, would we also have been prepared to judge that the wonderful consequences justify the means?

Again, these questions and the comparison they embody are likely to appear impertinent. Surely people are different from other animals, there is something special about human worth, human value, the dignity of human life. In light of that difference, that special something, we are entitled to do things to other animals that we should not do to people. Banting, Best and countless other scientists are morally justified in using nonhuman animals for medical research. The Nazi doctors who used human subjects in their "experiments" were moral monsters.

James Rachels' lucid, thoughtful and well-argued book leads us towards an uncomfortable re-evaluation of the simple distinction between people and other animals. *Created From Animals* ends where I have begun, with reflections on the use of animals in scientific research. The discovery of insulin is my case, not his, chosen to highlight some disturbing implications of the moral view that he develops. For Rachels' argument, grounded in an appeal to darwinian evolutionary theory, is that a simple reliance on the special worth of human beings cannot be sustained. If we are to draw a line between Banting and the Nazi doctors, then we must develop a more sophisticated moral view.

Rachels begins with an accessible account of Darwin's discoveries and a penetrating critique of some misguided proposals for relating ethics to evolutionary theory (primarily those of Herbert

Spencer and contemporary human sociobiologists). He then argues that Darwin undermined any substantial commitments in theology: consistent darwinians, Rachels claims, must abandon the central tenets of Christianity. Here, it seems to me, different audiences will be more or less challenged. Southern Baptists are likely to agree that the doctrines Rachels banishes are constitutive of "real religion"; liberal Anglicans are far less likely to be perturbed.

In any event, Rachels' main point is that darwinism undermines the doctrine that human beings have special worth and dignity by virtue of some human relationship to a deity.

The moral separation of humans and nonhumans must thus, he believes, depend on the differences between our species and others. But what kinds of difference might be morally relevant? Sensibly enough, Rachels focuses on sensitivity to pain and the capacity for intelligent behaviour. Appealing to studies of animal behaviour, he then builds a powerful case for claiming that the



Mary Evans

Soul mates? — Do humans "have special worth and dignity by virtue of some relationship to a deity"?

differences are matters of degree rather than differences in kind.

This leads directly to the final argument. Rachels proposes that we should be "moral individualists", justifying our ways of treating animals (whether human or non-human) on the basis of the characteristics that they have or lack, not on whether or not they belong to a particular group or species. It is thus as unjustifiable to defend "sacrificing" an animal on the grounds that it is not human, as it would be to justify lynching a man on the grounds that he belonged to a particular racial group. If the "sacrifice" is to be defended, then we must point out some relevant difference between the animal subject and human beings. Rachels suggests that the animal's inability to speak is not a morally relevant difference here — for we would not think it appropriate to "sacrifice" human subjects who were mute, or whose linguistic capacities had been destroyed.

Rachels leaves us with some conclusions, and with a challenge. Briefly surveying agricultural practices in the United States, he makes a strong case for the moral wrongness of factory farming methods, and argues that industrial testing of some products brings benefits too ephemeral to outweigh the suffering inflicted on the animals involved. Yet the principal challenge of the book, left unresolved in its final pages, is whether human beings are morally entitled to inflict pain and death on other animals even when there are actual or potential benefits for people. To return to the example with which I began: is it morally coherent to oppose the use of humans as experimental subjects (as, for example, in the Nazi camps), while regarding as justified Banting and Best's experiments on their unfortunate dogs?

There seem to be three lines along which an answer could be sought. One possibility is to scrutinize Rachels' contention that the psychological differences between humans and other animals do not amount to a morally relevant difference in kind. Perhaps our capacity for language and abstract thought brings with it the power to reflect on our own pain and impending death, adding an extra dimension to human suffering. A second approach, remaining within Rachels' framework, would insist that the differences of degree are reflected in the values of human and nonhuman lives. We could respond to the challenge by developing a consequentialist moral calculus. The values of the lives made possible by Banting and Best far outweigh those of the dogs who died prematurely. Finally, appealing to the principle that we can only be morally required to do what is possible for us, we might deny the psychological possibility of refraining from using experiments with animals as a means for alleviating the pain of members of our

own species. Could Banting, or anyone in his situation, have watched diabetic children die and resolved to save the dogs instead?

These are only indications of possible lines of response to Rachels' challenging reflections. *Created From Animals* leaves us with the task of working them out. Forceful, but never strident, it offers a moral perspective which thoughtful people, inside and outside the laboratory, would do well to ponder. □

Philip Kitcher is Professor of Philosophy at the Department of Philosophy, 0302, University of California at San Diego, La Jolla, California 92093, USA.

Flies from all angles

Gerald M. Rubin

Drosophila: A Laboratory Handbook. By Michael Ashburner. Cold Spring Harbor Laboratory Press: 1989. Pp.1,331. \$180.

Drosophila: A Laboratory Manual. By Michael Ashburner. Cold Spring Harbor Laboratory Press: 1989. Pp.434. \$50.

FOR over 80 years the fruitfly *Drosophila melanogaster* has been an experimental organism of major importance, first for genetics and more recently for cell, developmental and neurobiology. A vast literature of some 40,000 papers has accumulated along with much unpublished folklore. Newcomers to the field — several hundred graduate students and postdoctoral fellows each year — not to mention those of us with only a decade or two of experience, are understandably overwhelmed by this body of information and until now have had no practical way to gain access to it. Ashburner has performed an enormous service to the field by distilling nearly all the important facts and methods of *Drosophila* genetics into an encyclopaedia — *Drosophila: A Laboratory Handbook* — of some 1,300 pages. This book fills a long recognized void and has been enthusiastically received by the *Drosophila* research community.

Ashburner provides concise reviews of *Drosophila* nomenclature, genome structure, developmental biology and evolution in the opening and closing chapters. But the unique and most useful part of the handbook is the central 30 chapters in which the theory and practice of transmission genetics as applied to *Drosophila* are covered in a clear and comprehensive manner. When one considers the amount and breadth of the material covered — over 4,000 different primary sources are cited — it seems remarkable that it took Ashburner only nine years to produce the handbook. It is the very complete and

unbiased coverage of the literature, however, that leads to my one mild complaint. Little guidance is given to the reader as to which of the alternative approaches presented should be applied. Consider, for example, the chapter "Mutation and Mutagenesis". Spontaneous mutations as well as those induced by a variety of different methods, including ionizing and ultraviolet radiation, seven different chemical mutagens and viruses are all considered. The metabolism of mutagens, mutator genes, methods for administering mutagenic treatments, breeding schemes and statistical analyses are all discussed. The treatment is impressively comprehensive. If it has ever been done, Ashburner covers it and provides appropriate references for further reading. This extends to methods that most current fly workers are very unlikely to apply, such as vaginal douches to chemically mutagenize sperm. Although the advantages and disadvantages of each mutagen are mentioned in their respective sections, the novice would be greatly aided by a more general discussion of which mutagens and breeding schemes are the most suitable for each of the handful of applications commonly used in current work.

The companion volume *Drosophila: A Laboratory Manual* contains a collection of 137 protocols covering a wide variety of cytological and molecular techniques used with *Drosophila*. The protocols are clearly and concisely written and easy to follow. Protocols dealing with genetic methods such as how to administer common chemical mutagens would be a welcome addition. All protocol books, this included, suffer from the fact that active fields evolve rapidly so that many of these protocols will not age well and some are already out of date. Nevertheless, this is a useful volume and the appended lists of reagents and suppliers will be particularly valuable to anyone setting up a new laboratory.

These books are certainly an essential set of references for any laboratory working with *Drosophila* and I would be surprised if there were many fly labs that did not already own at least one copy. A shorter version of the handbook — say one-third of the length at one third of the price — would find a very wide market among individual students and postdoctoral fellows. This would be especially true if such a textbook of *Drosophila* genetics focused on that part of the handbook that is truly unique — the chapters on transmission genetics — and emphasized those aspects most relevant to current workers. □

Gerald M. Rubin is at the Howard Hughes Medical Institute and the Department of Molecular and Cell Biology, Division of Genetics, University of California, Berkeley, California 94720, USA.

Hart of the matter

W. F. Bynum

Mirror of Medicine. A History of the British Medical Journal. By P. W. J. Bartrip. *British Medical Journal and Clarendon Press: 1990. Pp. 352. £35.*

ROMANTIC philosophers and poets puzzled a good deal about whether the human mind should be likened more to a mirror or to a lamp. If to a mirror, it merely absorbed the external reality; if a lamp, it had the capacity somehow to impose order on that reality. The title of Peter Bartrip's fine history of the *British Medical Journal* uses the honourable but relatively passive image, but his narrative pinpoints several periods when the *Journal* has done a good deal more than simply reflect the medical knowledge and values of its time. Periodically, strong editorial leadership has transformed the *Journal* into an agent of professional and social change.

The occasion of Bartrip's monograph is the sesqui-centennial of the *Journal's* birth, although it was actually christened the *Provincial Medical and Surgical Journal* and was the entrepreneurial brain-child of its two initial editors, P. Hennis Green and Robert Streeten. Only gradually was its official relationship with what became the British Medical Association consolidated. The removal of the *Journal's* editorial offices to London was resisted by some members of the Association on the grounds that too much medical power was already concentrated in the capital; those who believed that neither the Association nor its *Journal* could otherwise become a truly national force eventually carried the day.

During its early decades, the *Journal* survived a succession of editors with variable capacities and commitment, but it was not until Ernest Hart's long reign, from 1867 to 1898, that the *Journal* became a genuine rival to its main weekly competitor, the *Lancet*, which had been established in 1823 by the equally colourful Thomas Wakley. Hart deserves a full biography, but at least Bartrip's sharp analysis brings him into clearer focus. His first stint in the editorial chair lasted a mere two years, followed by a terse resignation and disappearance for twelve

months from the medical scene, and possibly from Britain itself. His enemies (he had many) hinted that he had fled abroad to escape prosecution for the murder of, or at least culpability in the death of, his first wife. That he was reappointed so soon makes this unlikely. Whatever the reason, it hardly dampened his sense of self-importance: when the

tific and clinical papers it published. He championed Lister's surgical reforms, was committed to the role of experimental physiology and bacteriology in medical education and research, and advanced the claims of the profession as the natural guardians of the people's health. He could be an old-fashioned crusading journalist, as in his exposure of baby-farming practices. He gave public health and preventive medicine an ample hearing and if people sometimes worried that he had made the *Journal* more visible than its parent Association, this was probably not a bad thing, and, in any case, Hart was also an active BMA man.

Hart was a hard act to follow, but Bartrip also rates his successor, Dawson

Williams, as another of the *Journal's* great editors. He maintained high scientific and clinical standards in the papers he published and was particularly skilful in the way he managed things during the First World War. Dawson Williams' one major excursion into investigative journalism was the series of articles he commissioned on the contents of a large number of patent medicines. This resulted in a little volume, *Secret Remedies. What They Cost And What They Contain*, which became an edwardian best-seller and produced the inevitable sequel, *More Secret Remedies*. It also opened the stormy issue of advertisement in the *Journal*, since many of the pharmaceutical preparations, medical inventions and health aids advertised in its pages had dubious value. The BMA eventually established a committee to vet items deemed unsuitable for advertising in the *Journal's*



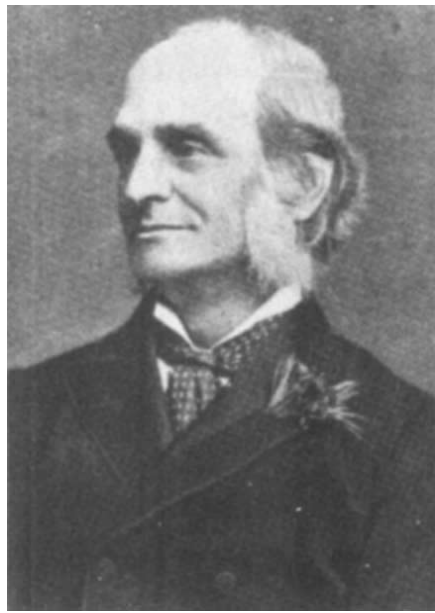
Doctor at large — Wandsworth Medical Treatment Centre in 1913, reproduced from *Mirror of Medicine*.

Journal (and Association) moved in 1887 to larger quarters in the Strand, Hart was unsuccessful in his bid to obtain the best room for his office, but he demanded, and got, a specially-constructed private entrance, complete with a plaque reading 'BRITISH MEDICAL JOURNAL, Editor'.

Overweening and overbearing he may have been ('scoundrel', 'vermin', 'beast', 'skunk' are some of the endearing names he was called), but his commitment to the *Journal* was absolute, even if his enemies might have claimed he took his cue from Louis XIV: '*Le British Medical Journal, c'est moi*'. Nevertheless, he turned the *Journal* around, in revenue, circulation and, above all, in the quality of the scien-

pages, but given the importance of this source of income, 'eternal vigilance' was not necessarily the committee's motto.

Dawson Williams dreaded retirement and consequently hung on long after his effectiveness had vanished. He was finally induced to retire, aged 73, in 1928, by which time the *Journal* was drifting without too much purpose. His successor, Gerald Horner, was already on the staff and was a great admirer of his boss. Unfortunately, Horner emulated only Dawson Williams' declining years. A weak man, Horner had trouble keeping good staff and surrendered editorial freedom to a management committee. Under him, the *Journal* was timid in taking up the important social issues of the interwar



Ernest Hart — a "sense of self-importance" (reproduced from *Mirror of Medicine*).

years, such as malnutrition, poverty and unemployment. Perhaps the most notable achievement of his period was the redesign of the *Journal's* format, by Stanley Morison, with the appearance of the now-familiar logo, the work of Eric Gill.

Horner was in ill-health during the last years of his editorship, and his deputy, Hugh Clegg, was effectively the editor for several years before being officially appointed in 1947, just in time to greet the birth of the National Health Service. Clegg was not enamoured of either Aneurin Bevan or his plans for the NHS, but he was realistic enough to recognize that medical opposition could not stop their implementation, and as the Appointed Day (7 July 1948) came nearer, he used the *Journal* to urge doctors to cooperate, and survived an attempt by some irate BMA members to have him sacked for his pains. Clegg proved to be a strong editor, as difficult personally as Hart, but equally effective in restoring the *Journal's* flagging reputation.

Bartrip ends his account with Clegg's retirement in 1965, although the present editor, Stephen Lock, has written a thoughtful postscript on the philosophy that has guided the *Journal's* production during the past quarter of a century.

Considering their importance, it is surprising that scientific and medical periodicals have attracted so little scholarly attention. Bartrip's monograph thus sets a standard and, one might dare to hope, a precedent. I can think of a number of other periodicals whose histories would make illuminating reading, of which *Nature* comes pretty close to the top of the list. □

W.F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Hypocrisies and half-truths

Roy Porter

Follies and Fallacies in Medicine. By Petr Skrabanek and James McCormick. Prometheus Books: 1990. Pp. 147. \$19.95.

THE recent inquiry, published in the *Lancet*, into the survival prospects of patients attending the Bristol Cancer Help Centre in the United Kingdom, elicited the most dreadfully predictable set of responses. Spokesman for orthodox oncology tended to say "we told you so — we always knew these quack treatments were worthless", while advocates of alternative medicine queried the statistics, or claimed their publication was "premature", because the result betrayed "hope". What is so refreshing, by contrast, about Skrabanek and McCormick's entertaining syllabus of medical errors is the independent gadfly spirit of the enterprise. Wielding the lancet of critical reason against nonsense, pomposity and illogic wherever they are found, the pair slaughter the sacred cows of all manners of medicine, regular and irregular, without fear or favour.

Alternative systems form, not surprisingly, the most vulnerable sitting targets, and the authors direct devastating shafts of wit against the more gullible sectors of the fringe. How can anyone truly believe that homeopathic medications work thanks to the extremes of dilution widely recommended (and mystifyingly dubbed 'potentization'), or that the essence of their efficacy lies in

the way they are shaken? Skrabanek and McCormick in their dossier of delusions dismiss acupuncture as "quackupuncture"; adjudge Christian science to be neither Christian nor scientific; and expose psychic surgery as a pernicious fraud.

But, when it comes to credulity, is orthodox medicine's track record so very much better? An entertaining chapter entitled "A Fistful of Fallacies" offers an alphabet of the absurdities of establishment medicine: from false faith in therapeutics (A recovers after taking medicine B, therefore B cures A: the causal link is all too often simply assumed), to semantic obfuscation and a terminal addiction to polysyllabics. Drawing upon Bacon's critique of the "idols of the mind", and adding a generous pinch of H. L. Mencken's short way with prejudices, the authors lay bare the sloppy handling of statistics still pervading so many supposedly scientific clinical trials, and expose the way dubious diagnostic labels all too often serve as cloaks for ignorance and surrogates for understanding.

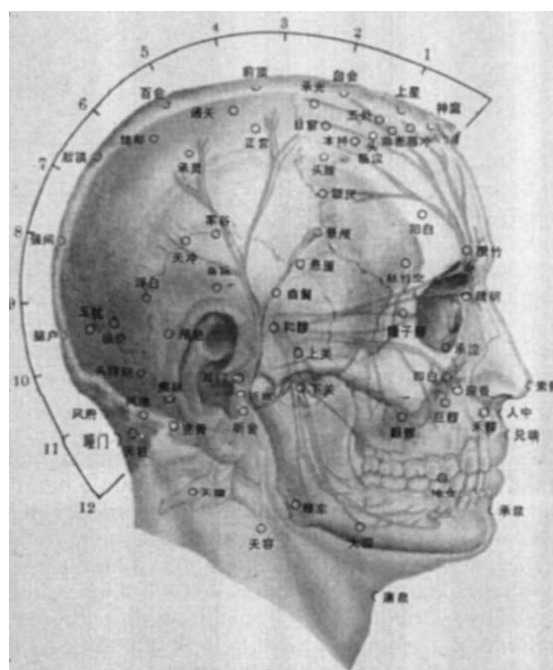
They are not critical of medicine *per se*, Skrabanek and McCormick insist, they merely want criticism in medicine. Nor is their stance wholly destructive. On the positive side, they argue that because medicine cannot be an exact science — it is, rather, a technique of healing — the human factor must be paramount.

Hence they have some robust recommendations about healing. All surveys confirm that the doctor can wield a magical power. Inert drugs, prescribed with the right flourish of confidence, produce a placebo effect; the bedside manner is itself therapeutically effective. Why, then — prejudice aside — has so little attention

been paid to the promotion of this kind of faith-healing within medicine? Isn't it time medical education capitalized upon the obvious healing power of the placebo?

'Prevention' forms another instance where specious theory may be taking precedence over the real interests of the patient. In the abstract, the extension of screening procedures (for example, for cervical cancer) may seem incontestable. But, taken in the round, do the outcomes actually bear this out? Include all the undesirable, and, too often, ignored, side-effects — false positives, surgical intrusions and the vast and mainly needless psychological anxiety created — and it is less clear where the balance of benefit lies.

Avoiding cynicism or



Acupuncture or "quackupuncture"?

Science Photo Library

nihilism, Skrabanek and McCormick demand constant critical vigilance and radical scrutiny of the taken-for-granted. Are we not losing our sense of proportion if we assume the yard-stick of medical progress is the prolongation of life at all costs? May not quality of life, freedom from anxiety and individual liberty count for more? It is fashionable to scare people into health with campaigns against cholesterol and the promotion of regular check-ups; but exposing people to half-baked diets and additional doctoring may be just as harmful.

Occasionally, Skrabanek and McCormick hoist themselves on their own petard, when their debunking passion — evidently learned from Martin Gardner — gets the better of their own critical faculties. They assert, sweepingly, that “the present activities of surgeons general, health education councils, and many

academic departments of public health and the like, are in danger of corrupting medicine by morality”. Maybe, but, by their own criteria, we’re at least entitled to some evidence of this. All the reader gets, however, is an anecdote from Manila and a quotation from Plato: hardly proof positive of a point whose rhetoric is rousing but which actually requires serious evaluation.

Yet *Follies and Fallacies* never expects to be the last word. Fired especially at medical students, it aims to provoke more than to prove. And in this it succeeds, and leaves one looking forward to a further training of the Skrabanek and McCormick artillery against hypocrisies and half-truths. □

Roy Porter is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Hominid whodunnit

G. Ainsworth Harrison

Piltdown. A Scientific Forgery. By Frank Spencer. Oxford University Press: 1990. Pp. 272. £17.95, \$24.95.

The Piltdown Papers. By Frank Spencer. Oxford University Press: 1990. Pp. 282. £30, \$65.

THE fascination of the Piltdown forgery is apparently endless. It was indeed one of the most gigantic frauds in scientific history: it bedevilled understanding of the fossil record of human evolution and went unexposed for more than forty years. Perhaps the continuing fascination comes from the fact that the circumstances of the perpetration of the fraud remain obscure, and the guilty party or parties not unequivocally identified. Whether one can get nearer to the truth eighty years on than J. S. Weiner’s exposure of the fraud in 1953 is rather doubtful, but nonetheless Spencer’s book is a scholarly and important contribution to the now enormous literature dealing with the Piltdown forgery.

The nature of the original finds, which were apparently extracted from the lower Pleistocene gravels near Uckfield in Sussex, in the United Kingdom, between 1912 and 1915, are well known: the finds consisted of substantial parts of a human cranium which was essentially modern; much of a mandibular horizontal ramus which was very ape-like; a canine-tooth which fitted perfectly the prediction of Sir Arthur Smith Woodward; sufficient of a second individual to show that one was not dealing with a fortuitous association; and some artefacts and fossil mammals. All turned out to have been ‘salted’, and the

jaw and teeth ‘doctored’ to show some human characters.

Subsequent fossil discoveries, particularly in Africa, made Piltdown ever more anomalous. Human evolution proceeded by modification of the jaws before substantial change in the size of the brain, rather than the other way round. Nonetheless, authors of books on palaeoanthropology in the 1940s still gave a central place to Piltdown, often identifying it with the ancestry of modern man. In such schemes, other fossils were all side branches. That such a position could be sustained for so long is remarkable — at least with hindsight! It was not helped by the fact that almost everyone working on

Piltdown had to work with caste material. But eventually Weiner recognized what a nonsense it was, proposed the forgery explanation, demonstrated in collaboration with K.P. Oakley and Sir Wilfrid Le Gros Clark the scientific basis to the fraud and undertook his own investigation into the perpetrators. He concluded in his book *The Piltdown Forgery* (1955) that Charles Dawson, a local solicitor, amateur naturalist and the ‘discoverer’ of the first finds, almost certainly had to be heavily involved; but whether he was alone or had accomplices was not completely clear. On balance, Weiner favoured Dawson’s sole involvement.

Since then many others have tried their detective skills, and fingers have been pointed at almost every possible figure associated with the finds or with Dawson. It seems that one only needed to have purchased potassium dichromate (which was used to ‘fossilize’ the bones) around 1912 to have come under suspicion. Strangely, perhaps, one of the few people to escape accusation has been Sir Arthur Keith, then conservator of the Hunterian Museum of the Royal College of Surgeons, who devoted much of his life to researching Piltdown, particularly the reconstruction of the cranium from the fragments recovered. That omission is corrected in Spencer’s book, in which the proposition is developed that Keith was Dawson’s scientific accomplice and perhaps instigator of the whole affair.

Keith’s involvement was first seriously suggested by Ian Langham, a young Australian historian and anthropologist who died tragically in 1984. Langham’s work



The ‘Piltdown gang’ as painted by John Cooke — Arthur Keith is seated (centre) wearing a labcoat, Charles Dawson and Arthur Smith Woodward are standing (left and right, respectively) to the right of the painting.

was passed by his family to Spencer to develop and publish. Spencer has not only done this but also set it within the whole Piltdown saga. He clearly shares the same verdict of guilt as Langham. But their case is not only circumstantial, as it is almost bound to be, but also, in my opinion, very thin. It largely rests on Keith apparently knowing a little more about the circumstances of the find than might initially have been expected — and some of this assertion depends on a judgement of what might have been conveyed in an hour's conversation between Smith-Woodward and Keith. There is also a suggestion that Dawson and Keith may have known each other rather better around the time of the discoveries than they publicly acknowledged.

I have to say I am quite unconvinced by this evidence. I am, however, privileged to know J. S. Weiner's views on Keith's reaction to the fraud's exposure. I was in effect acting as a research assistant to Weiner at this time, and accompanied him on many occasions during his investigation. Unfortunately, I was not present when he interviewed Keith, but I remember his account the following day. He said that Keith appeared utterly taken aback and dismayed by the idea that Piltdown

was fraudulent. He had no doubt at all that Keith was not involved.

In the second edition of *The Antiquity of Man* published in 1925, Keith devotes 248 out of 734 pages to Piltdown. By this time he had received a knighthood and was a Fellow of the Royal Society. Would he by then have spent so much time and effort and identified himself so closely with something that he knew was false? Why did he have so much difficulty reconstructing the cranium if he had been party to the construction of the fake? For me the accusation just does not ring true. I do, however, applaud Spencer's book as a whole, which is well researched and attractively written. With the accompanying volume of collected papers and letters it provides the definitive history of Piltdown: the 'discovery', the exposure and the long search for the culprit. Spencer also documents not only how damaging this fraud was to science, but also to the reputations of those who were unfortunate enough to be in any way associated with it. □

G. Ainsworth Harrison is in the Department of Biological Anthropology, University of Oxford, 58 Banbury Road, Oxford OX2 6QS, UK.

Integrity on trial

Ralph A. Lewin

Where The Truth Lies. By Jan Sapp. Cambridge University Press: 1990. Pp.340. Hbk £30, \$59.50; pbk £15, \$18.95.

JAN Sapp has reviewed the life and times of Franz Moewus, and discussed implications of his scientific publications. If you have heard of Moewus (1908–59) you will certainly want to read Sapp's review; if you have not, I urge you to do so anyway. The Moewus saga (a saga is a story, not always completely truthful) is wider in scope, in time and in moral implications than those of Piltdown Man or Kammerer's black-toed toads. Was his work a tissue woven largely of lies, or a set of insightful if irreproducible experiments? Was Moewus a fraud or a prophet, or a bit of both? He was certainly bright — but then, an embezzler has to be bright if he is to succeed in beating the system.

The Moewus story began in Germany, interluded in Australia, and ended in the United States. Most of his publications dealt with a hitherto obscure little alga, *Chlamydomonas*, but they had implications for our understanding of microbial sexuality, physiology, biochemistry and genetics. Indeed, in 1940 many considered that Moewus' work was at the forefront of microbial biochemical genetics (if not

actually of molecular biology, which I think started somewhat later). He was backed by powerful scientific figures: M. Hartmann and R. Kuhn in Germany, and later Sonneborn in the United States, although almost everything he reported was at best irreproducible, and at worst downright fishy. His final scientific demise (as Sapp puts it) was in a dénouement at the Marine Biological Laboratory, Woods Hole in 1954.

Should he be dismissed simply as a charlatan? Sapp, a science historian and philosopher, thinks not, for at least two reasons. Moewus, he contends, saw the light and pointed the way to microbial genetics as soon as, if not sooner than, the Nobel laureates G. Beadle, E. Tatum and J. Lederberg. If he fabricated or fiddled his data — and there is irrefutable evidence for this — his peccadillos could be partly excused by the political and social systems in which he worked. All scientists do a little fiddling, Sapp avers; indeed, some fiddling is almost expected of all of us, and in Nazi Germany this was more true than elsewhere.

I have read Sapp's book twice, and I cannot fully agree with his premises. He reiterates that there is more truth in fiction than in fact — but is there? Most scientists I know plod along, year by year, grant by grant, finding out little items about the nature of things and reporting them reasonably objectively (that is, truthfully). Few of us (I can certainly speak for us biologists) espouse doctrines so enthusiastically that we adjust our data to fit

them. (Admittedly there was the nasty aberration of lisenkoism, but surely that was a historic exception.) Few of us belong to specific schools of dogmatic thought, as Sapp would have us believe, publishing our results as adversarian treatises, as protagonists of this doctrine or antagonists of that. (This may be true among clergymen, philosophers, and perhaps to lesser degrees among linguists or economists, but not us biologists.) If Sapp had been a scientist himself, I think he would have agreed with me. But he is not, and he argues forcefully and repetitiously that truth is mostly in the mind of the scientist, as is beauty in the eye of the beholder.

Sapp has done an extensive piece of archival research. This book deals with some 70 publications by Moewus and some 40 more by his coauthors, his reviewers and his detractors. (Certain relevant publications by L. Wiese, K. Thimann and M. Hagen-Seyffert have received less attention than they merit in this connection.) It is also based on extensive interviews with various people who knew Moewus and who have ideas about his publications — including Mrs Kobb, Moewus' widow — and on almost 90 items of correspondence and other peripheral literature. (A contemporary and compatriot of Moewus' to whom I showed Kobb's paragraphs dismissed much of her pleading as "pathetic nonsense", and suggested that Sapp might have been more critical.) Some 50 of the latter items were from the extensive files of the late T. M. Sonneborn, initially Moewus' strongest supporter in America and, later, one of his most disillusioned correspondents. Sapp has also added chapters embodying his own ideas about truth in science and the way it is elucidated, about the work of Gregor Mendel (some of whose data were also statistically questionable, though the orders of magnitude of the respective improbabilities were considerably different), and about the general origins of microbial genetics. The relevance of some of this is peripheral, but it is certainly of educational value.

I think all kinds of scientists would profit by reading this book. But I hope it does not fall into the hands of scientifically illiterate pundits in law or politics; they might get wrong ideas about the integrity and veracity of scientists in general. After all, truth will out in the long run. If you question this, then look into the latest compilation of scientific information about *Chlamydomonas*, edited by Libby Harris in 1989, and see how little credence has eventually been placed in the publication of Moewus. □

Ralph A. Lewin is at the Scripps Institution of Oceanography, University of California, La Jolla, California 92093, USA.

Semidetached observation

W. C. McGrew

Through a Window. My Thirty years with the Chimpanzees of Gombe. By Jane Goodall. Weidenfeld and Nicolson: 1990. Pp.229. £15.00, \$21.95.

SUCH is the appeal of field primatology that its better-known practitioners have been able to establish a pattern of parallel publication. That is, while producing the obligatory scholarly books and journal articles, they also often write handsome and engrossing counterparts for the wider audience of the educated public. The benchmarks for this dual-channel communication were George Schaller's *The Mountain Gorilla* and *The Year of the Gorilla*, respectively. Now, after giving us her scientific opus, *The Chimpanzees of Gombe* (1986), Jane Goodall has taken the process a step further with *Through a Window*. The key is in the subtitle, which identifies it as the sequel to her earlier *In the Shadow of Man* (1971, reissued, 1989). The resulting volume matches the high quality of its predecessor and should prove to be just as popular. (Remarkably, *Shadow* has been translated into 48 languages. Can any other book of natural history match this?)

In the Shadow of Man recounted the first ten years of Jane Goodall's study of the behaviour of a small population of wild chimpanzees living on the eastern shore of Lake Tanganyika. It chronicled the foundation and rapid rise of the Gombe Stream Research Centre through the 1960s. *Through a Window* carries on for the next 20 years, through the 1970s and 1980s. The formats of the two books are virtually identical: around 100,000 words of text divided into 20 punchy chapters, illustrated by about 100 excellent photographs, of which roughly a fifth are in colour. Each book is good value for money. Each has a compulsively engaging opening chapter, and a conscience-rousing closing chapter. (In *Shadow* the latter was called "Man's Inhumanity", whereas in *Window* it is called "Our Shame".) The linkages between chapters are tight, so that they unfold episodically, like a serial.

But there are differences between *Shadow* and *Window*, both in format and content. Neither has tables or graphs, but the sequel abandons pedagogy altogether: references, systematic appendices and the useful genealogical chart are omitted from *Window*, as are David Bygott's splendid drawings. Two aspects reflect the personalized focus of the newer book: the index uses helpful codes to distinguish human from ape characters, thus making clear that David Greybeard is a Gombe chimpanzee subject, whereas David Riss is a human student. Similarly, a third

of *Window*'s chapters are individual case histories and are labelled as such, whereas none of *Shadow*'s chapters are so constructed.

So, what have 20 further years of fieldwork at Gombe revealed about chimpanzees? As Goodall has repeatedly stressed, life-long study of a long-lived creature continues to yield new findings. The 1970s showed a darker side of chimpanzee nature that counter-balanced the idealized positive picture from earlier years.



Monkey see, monkey do — Young chimpanzees learn much of their social behaviour by watching adults. This one is fascinated by his mother being groomed. (Reproduced from *Through a Window*.)

The males of Gombe's larger, Kasakela community relentlessly exterminated the smaller, neighbouring Kahama community, and in the process usurped its territory and females. The details of this "four-year war" make gruesome reading. Equally surprising and sobering was the appearance of the killing and eating of infant chimpanzees. Whether these are effective strategies that enhance the reproductive success of their perpetrators, or are pathological responses to overcrowding cannot yet be determined, but similar events have occurred in the other, bigger population of chimpanzees in Tanzania, as studied in the Mahale Mountains, by Toshisada Nishida and his colleagues.

In contrast, one of *Window*'s most uplifting chapters documents several cases of successful adoptions of orphaned youngsters, either by a relative or a 'family friend'. These foster care-givers may invest much time and effort that make a life-and-death difference to the recipients. Only continuing longitudinal study will

show if this apparent altruism is eventually rewarded or not.

Goodall devotes less space to Gombe's history, but one event stands out as crucial. In May 1975, guerillas from neighbouring Zaire kidnapped four expatriate researchers and held them for ransom. Eventually their release was achieved, but the implications for security were grave enough to alter the norms for research at Gombe. Tanzanian field assistants took on the burden of behavioural observation, replacing Western students. Continuity of data collection was thus maintained, but exceptional research efforts such as the marathon watch on Gombe's dominant male were lost. (In 1974, two Stanford undergraduates, Curt Busse and David Riss, did 50 days of consecutive dawn-to-dusk follows of Figan, the alpha male of the Kasakela community. This is arguably

the single most impressive feat of field observation ever.)

It is inevitable that 30 years of research raises some question marks of methodology. "Provisioning" of the chimpanzees with prized but unnatural goods like bananas long ago dropped to trivial levels, but intervention continues. Sometimes, this has shifted from reactive to preventive: when twins were born, their mother Melissa "... was followed every day, for we all feared that Passion and Pom [the main infant-killers] would strike again and we planned to intervene if they did". Such management of the lives of wild subjects is bound to be controversial for scientific purists, but Goodall has never shrunk from her stated position as more than a detached observer. This accords with the greater emphasis in the sequel given over to chimpanzee welfare in captivity. Goodall squares up to this theme in an appendix on the exploitation of animals by laboratory scientists. Her use of terms like "brainwashing", "self-serving" and

"torture" will guarantee that these topical issues will be scrutinized by readers.

In a word, *Window* is a very personal account, on all fronts. The personalities of individual apes are as carefully presented as any cast of characters in a play. Conversations with colleagues, quoted verbatim, convey the excitement of collaborative research in the bush. Again and again, telling vignettes make the point more compellingly than any judiciously qualified generalization could do.

Most importantly, Jane Goodall's viewpoint (and so her personality) have virtually defined our impressions of chimpanzee nature, at least in the English-speaking world. This is an awesome responsibility for any scientist.

When Louis Leakey arranged for Jane Goodall to begin studies of the chimpanzees at Gombe, he predicted that it would take ten years to understand them. This seemed excessively cautious at the time. Now, after 30 years, one can only look forward to the next decade of continuing research, and to Goodall's next account of it. □

W. C. McGrew is a Reader in the Department of Psychology, University of Stirling, Stirling FK9 4LA, UK.

It's a jungle

Tim Lincoln

Brazzaville Beach. By William Boyd. Sinclair-Stevenson, London: 1990. £13.95. To be published in the United States early next summer by Morrow.

IN this, his fifth novel, William Boyd has returned to home ground, to Africa, for one of the two intertwined stories that make up the narrative. But otherwise the book is a daring and imaginative departure from his previous works. The principal character is a woman, the young and self-assured Hope Clearwater, a botanist turned ethologist, into whose skin Boyd has convincingly written himself (convincingly, that is, as far as a male reader can tell). And the subject matter is mathematics and primatology, with at the bottom of it all the moral that in intellectual endeavours winning, or being seen to win, is everything.

The first story centres on the shadowy figure of John Clearwater, said to be a brilliant mathematician. After four years at Caltech working on game theory, he has returned to Britain, to Imperial College in London, to meet and marry Hope, and to match his brains against the challenge of discovering a simple formulation for turbulence. The second story (chronologically the later) is set in a vaguely defined, war-ravaged part of Africa, at the Grosso Arvore primate research centre. It

is there that Hope has gone, after the mental battering unwittingly meted out by her husband, to work under Eugene Mallabar, who has devoted the best part of his life to the study of chimpanzees. Interspersed between the two are brief ruminations, which veer close to the extremes of pretentiousness and profundity, and which in large part come from the present, from Brazzaville Beach, where Hope has eventually washed up. From the



Jane Goodall — inspiration for Hope? (Reproduced from *Through a Window**).

beach she examines her life and muses on why her marriage to John Clearwater, and the research at Grosso Arvore, went so badly awry.

Both episodes are narrated in spare and beautifully controlled style. John Clearwater slides into insanity, inch by inch, by way of the fevered digging of holes in search of inspiration, an affair with the wife of an acned physicist, and (towards the end) treatment by electroconvulsive therapy. His agony is to have been on the verge of, but to have been beaten to, what would have been a claim to mathematical immortality in the form of the Clearwater set — a simple algorithm "that would reproduce the magical, infinite variety of the natural world". Life amongst some of the bit players is none too happy either. There is for example Hope's gruesome sister, married to a solicitor and "sinking in the quicksand of prudence, moderation and propriety" in the stuffy respectability of the English home counties.

But it is in Africa, among the chimpanzees, that the starkest of horrors (and the best joke) lie. To followers of the career of Jane Goodall* the chain of events will be

* Jane Goodall's most recent book, *Through a Window. My Thirty Years with the Chimpanzees of Gombe*, is reviewed on page 371 of this issue.

uncannily familiar. Boyd has founded his fiction curiously close to fact, in that for Grosso Arvore one could well read Gombe — not only in the background (research students not merely from the United States but explicitly from Stanford, the evening chore of writing up field notes, the controversial existence of an artificial feeding area, the round of lecture tours), but in the two main occurrences. Of these one is a kidnapping of researchers; the other is a north-south rift in the chimpanzee community, subsequent brutal attacks by the northerners upon the break-away group, and the shattering observation that chimps are capable of infanticide and cannibalism.

At this point, Boyd parts company with Goodall. In discovering the 'chimpanzee wars', Hope threatens Mallabar's cherished theories, his sources of money and his celebrity status, built in part through his books *The Peaceful Primate* and *Primate's Progress*. But experienced operator that he is, the shocking revelations are deftly appropriated as his own. When the third book appears, Hope, like her former husband, has become a loser, literally a footnote to history.

Brazzaville Beach is probably best taken as no more and no less than a compelling novel. Yet, together with other straws in the wind, it is tempting to see in it a sign that one of the two cultures (a matter now widely and wrongly considered *passé*) is taking the other seriously. That hope is no doubt unfounded. There is, though, the other Hope — William Boyd should persuade himself to get his splendid literary creation off the beach and back into the scientific jungle where she belongs. □

Tim Lincoln is News and Views Editor of *Nature*.

Original lingo?

John C. Marshall

Language and Species. by Derek Bickerton. Chicago University Press: 1990. Pp. 297. \$24.95.

BY the close of 1832, the *Beagle* had reached Tierra del Fuego and was anchored in Good Success Bay. Captain Fitzroy sent a party ashore, and the young Charles Darwin met his first Fuegians in their native habitat. In his diary for 18 December, Darwin wrote: "Their language does not deserve to be called articulate. Capt. Cook says it is like a man clearing his throat; to which may be added another very hoarse man trying to shout and a third encouraging a horse with that peculiar noise which is made in one side of the mouth." Some forty years later, Darwin had the insight (and the courage)

to admit he was wrong ("... I took a very erroneous view of the nature and capabilities of the Fuegians").

Nonetheless, the diary entry does reveal why it is nowadays so acutely embarrassing to read most of the anthropological linguistics written in the nineteenth century. Language is what makes us human, but some are more human than others. Colonial exploitation (economic, cultural and religious) demanded that, in the great chain of being, 'primitive savages' should fall roughly half way between chimpanzee and bourgeoisie. Exponents of the theory of 'mental evolution' duly interpreted their observations to show that the speech of 'lower races' was mere jabbering, eked out by pantomime, and that their mentation consequently allowed neither abstraction nor generalization.

In the twentieth century, two basic facts put the damper on speculation about language origins. First, all extant languages have equal expressive power (although achieved by superficially distinct means); and second, the fossil record of the hominid line will not support reliable argument about the cognitive capacities of the brains that once inhabited those skulls. For the most part, the domain became the province of the dilettante. It must accordingly have required considerable chutzpa for a professional linguist to launch wholeheartedly into *Language and Species* as an earnest topic of extended research.

Derek Bickerton is professor of linguistics at the University of Hawaii. Much of his own fieldwork has been on pidgins (which may in some sense be 'primitive' or protolanguages) and their conversion into creoles (which are definitely 'the real thing'). A creole language "is what results when a pidgin, created by adults, is learned by the children of those adults." The interest of the transition lies in the observation that whereas pidgins are pretty much a jumble of words, creoles exhibit "the same type of structure as any other natural human language". The propositional meaning of an utterance in a pidgin must be 'guessed' on the basis of lexical semantics and the context of situation; but the determination of 'thematic roles' (who did what to whom) in a creole sentence is controlled by the formal principles and parameters of universal grammar.

If, as Bickerton (coherently, albeit controversially) argues, the "jump from protolanguage to language (can be made) in a single generation", the biological implications are strong and profound: the process of creolization becomes a laboratory in which we can see (in relatively 'pure' form) the innate predispositions that the child brings to language acquisition as a species-specific characteristic. The behaviourist alternative to 'nativism' — that a general purpose learning device 'induces' grammatical structure from a

rich input of exemplary instances — becomes decidedly unappealing if the features acquired are simply not there in the pidgin environment. In Bickerton's view, creoles are thus "an unusually direct expression (of the) capacity to recreate language in the absence of any specific model from which the properties of language could be 'learned' in the ways we normally learn things."

Language and Species sets this discussion in the context of Noam Chomsky's government and binding theory (the only serious account of syntactic structures that currently exists). Bickerton's lucid description of the theory serves to draw a fairly clear line between protolanguage and the full system. That demarcation permits Bickerton (with justification) to have his cake and eat it with respect to the evolution of language. The gestural devices that can be taught to apes, the 'language' of small children (under the age of two), and the 'invented' pidgins that allow communication between adult speakers of mutually incomprehensible languages seem to fall together as limited, yet extremely useful systems of representation and expression. But there is a qualitative Rubicon between that level of linguistic labelling and the formal syntactic properties that then emerge in the language of the older child and that characterize all natural languages, creoles included.

Bickerton can thus avoid the counterintuitive view that the capacity for language in the species developed wholly out of the blue whilst recognizing the abruptness of the transition to true language. That the notion of 'gradualism' in both the phylogenetic and the ontogenetic development of language has been a conspicuous failure does not imply that there is no functional precursor to act as a springboard to syntax. Yet Bickerton's "language biogram hypothesis", derived from his creole data, is actually a more extreme 'nativist' position than that adopted by Chomsky. Chomsky's 'selectionist' theory of language acquisition presupposes that the child is presented with the linguistic evidence required to set a parameter of the syntax to the value represented in the adult language to which he or she is exposed; but Bickerton seems to be claiming that creoles possess features for which there is literally no evidence in the linguistic environment. Concerning the neurobiology of such leaps, however, Bickerton can add little to the handwaving that has always been such a lamentable feature of attempts to go beyond description in this area.

It would be unfair though to conclude on that kind of negative note. The evolution of language is a fascinating topic, and insofar as we know anything about it, Bickerton's *Language and Species* is the best introduction we have. On the other

hand, Bickerton himself stresses the pessimistic, sociocultural reflection that language enables us to create a wide range of "dysfunctional representations" of human nature which may yet kill us all. It was not the apple that tempted Eve so much as the seductive sentiments of the serpent. □

John C Marshall is at the Neuropsychology Unit, University Department of Clinical Neurology, The Radcliffe Infirmary, Oxford OX2 6HE, UK.

Purplish prose

Henry Gee

Lucy's Child: The Discovery of a Human Ancestor. By Donald Johanson and James Shreeve. Viking: 1990. Pp.318 £16.99. Published in the United States by Avon, New York. \$10.95 (pbk).

SCHOOLCHILDREN will be familiar with the class of creative writing generally described as "what I did in my summer holidays". The idea behind this literary form is to convince the reader that one has spent the vacation engaged in various improving pursuits, rather than hanging out on street corners, following the exploits of the Teenage Mutant Ninja Turtles or listening to Bon Jovi records. Typical examples of the genre inflate every routine trip to the shops into an epic voyage of discovery.

Donald Johanson had hoped to spend his summer holidays exploring palaeontological virgin territory in Ethiopia in search of fossil hominids to add to "Lucy" and the so-called "first family" — the spectacular discoveries that formed the basis of his reputation. A sequel to his bestselling *Lucy: The Beginnings of Humankind* would surely have been on the cards. But thwarted by the tides of politics, he had to make do with the well-worked and tourist-encrusted rocks of Olduvai Gorge in Tanzania instead. A fine shopping expedition was beginning to take shape.

Nevertheless, Johanson and a small crew unearthed remains of a diminutive hominid in sediments dated at 1.8 million years old. The description was eventually published in *Nature* (327, 205–209; 1987), which is not bad for a holiday jaunt. The hominid was catalogued as OH62 (short for Olduvai hominid number 62) and was interpreted in the paper as *Homo habilis* despite its very small stature and postcranial features arguably reminiscent of *Australopithecus*. To attribute OH62 as *Homo* put further strains on the validity of *H. habilis* as a single taxon (see the News and Views article by Bernard Wood which accompanied Johanson's paper). Naturally enough, much is made in *Lucy's Child*

of the specimen's attribution to *Homo*, the competing arguments, and What It All Might Mean.

But if abstruse anatomical arguments are not to your taste, *Lucy's Child* has something for everyone: there are florid descriptions of what it is like to be involved in a field season, vignettes of street life in Dar-es-Salaam and effusive pen-portraits of many of the scientists involved in the dig. The abundant flashbacks to critical moments in our understanding of human evolution framed by purplish descriptions of East African sunsets add to the impression of an uneventful amble to the drugstore made into a ten-day hike across the Siberian tundra pursued by ravening hordes of mammoths. The story starts, even, with a situation comedy: Johansen stuck in the sweltering night of his hotel in Dar-es-Salaam, paradoxically freezing cold because of a badly

adjusted air conditioning unit.

This baby-with-bathwater approach, though, falls into the classic trap of the summer holiday genre. Unless the reader is engaged right from the beginning, all these distracting details will only raise suspicions that there really is not much to write about after all. TV dinners that look so colourful tend to be, in the event, profoundly unsatisfying, and the shopping trip was, after all, just a shopping trip. Here, as with many sequels, the good bits were all done the first time round, and the rest is dressed up with special effects. Which is a pity, because, with all its faults, *Lucy's Child* is a cracking good read. Just the sort of thing to while away lazy street-corner afternoons. But not really serious enough when it's time to put away the skateboard and head back to school. □

Henry Gee is on the editorial staff of *Nature*.

The smallest continent

Jared Diamond

Mammals of New Guinea. By Timothy Flannery. *The Australian Museum (Robert Brown and Associates): 1990. Pp. 440. A\$79.95.*

The Birds of Papua New Guinea. By Brian J. Coates. *Dove Publications (Box 59, Alderly, Queensland 4051, Australia). Volume 1 (Nonpasserines): 1985. Pp. 464. US \$85, A\$110, £60. Volume II (Passerines): 1990. Pp. 576. US \$110, A\$140, £70.*

AUSTRALIA is famous for having the world's most distinctive mammal fauna, described and illustrated in numerous books. No text of evolution or comparative physiology omits examples based on those mammals. But few people are aware that the nearby minicontinent of New Guinea, with only one ninth of Australia's area, has a related but almost as distinctive and rich mammal fauna. The reason for this unawareness is simple: no comprehensive account of New Guinea's mammals has been available. That deficiency has now been filled by Flannery's book.

The book begins with brief accounts of New Guinea geology, climate, vegetation, zoogeography and extinct mammals, which included cow-sized marsupial herbivores. There follow detailed accounts of all New Guinea's 187 known extant native mammal species, including colour photographs and new field information obtained by the author. Apart from two echidna species, the mammals comprise nearly equal numbers of marsupials, rodents and bats, each group having undergone multiple radiations within New Guinea. Compared to Australia and

southeast Asia, New Guinea is notably rich in frugivorous bats, aquatic insectivores, and small arboreal folivores, and notably deficient in carnivores (only five species, none weighing over 2 kg). The differences from Australia may reflect New Guinea's smaller size, tropical location, much wetter climate and glacier-capped mountains up to 5,000 metres high. The differences from Asia reflect the inability of flightless placentals other than rodents to reach New Guinea, and the resulting radiation of New Guinea mammals to fill niches occupied in Asia by placentals such as primates and squirrels.

Mention of some New Guinea specialities described by Flannery will help convey a sense of his book. Among New Guinea's marsupials are: six species of large kangaroos that live in trees instead of on the ground; phalangiers and ringtails, the marsupial equivalent of folivorous primates; and the possum-like *Dactylopsila palpator*, strikingly convergent on Madagascar's aye-aye in possessing a very long finger used to pull grubs from rotten logs, and convergent on American skunks in its smell and bold black and white pattern. The murid rodent fauna, one of the world's richest and most highly endemic, includes giant rats weighing up to 1.7 kg, many endemic genera of herbivores and an earless rat confined to cold mountain torrents where it somehow manages to conserve body heat, catch insects and not get swept away. New Guinea's bats include one of the world's largest (a *Pteropus* flying fox with a 1.5-metre wingspan), the highly specialized and brightly coloured subfamily Nyctimeninae (unusual among Megachiroptera in being insectivores), and the tragic Bulmer's fruit bat (formerly believed extinct for 9,000 years, rediscovered alive in abundance in 1975, and re-extirminated by hunters two years later).

Although Flannery's book will be indis-

pensable to vertebrate biologists studying the Australian region, it will find a much wider audience among evolutionary biologists, ecologists, anatomists, physiologists and sociobiologists, for whom this previously inaccessible fauna extends our knowledge of the limits of mammalian adaptation.

New Guinea's endemic birds are as remarkable as its mammals and include birds of paradise, bowerbirds, mound-builders and a marsupial-like radiation of songbirds. Unlike the case for New Guinea's mammals, its birds are often cited in general biology texts and have been well served by two good modern books, the field guide by B. Beehler *et al.* (*Birds of New Guinea*: Princeton University Press; 1986) and the out-of-print handbook by A.L. Rand and E.T. Gilliard (*Handbook of New Guinea Birds*: Weidenfeld and Nicolson; 1967). Coates' two-volume treatment, of which volume 2 has just been published, nevertheless makes important new contributions. By focusing on the nation of Papua New Guinea, it includes the birds of the Bismarck Archipelago east of New Guinea, for which no modern account has previously been available. It also gathers together for the first time in book form extensive information on the behaviour of all New Guinea bird species.

But for biologists in general the most notable feature of Coates' volumes is the huge collection of colour photographs, which to my prejudiced eye includes some of the most stunning and instructive pictures of animals ever published. Especially valuable are Coates' photos of displaying birds of paradise and bowerbirds, which have been coming to the fore in recent discussions of sexual selection. Most biologists are now aware that the former have the most elaborate sexual display plumage among birds, and that the courtship bowers of the latter are the most elaborately decorated structures erected by any animals except humans. Yet the widely cited verbal descriptions are no match for these revealing photos, which will make concrete for sociobiologists the behaviours whose costs and benefits they now debate. At the risk of singling out five of Coates' photos for special mention, I suggest readers turn first to the masochistic display of Archbold's bowerbird; the belly of the twelve-wired bird of paradise — like a bright yellow light bulb carried below the bird in flight; the fluorescent blue tail wires of the magnificent bird of paradise perched over the display court that it has cleared; and the emperor and blue birds of paradise spreading their plumes while hanging upside-down from a branch. □

Jared Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.